CHARACTERIZATION OF THE CHEDIAK-HIGASHI SYNDROME GENE IN HUMAN AND MOUSE

By

VELIZAR T. TCHERNEV

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1998

Copyright 1998

by

Velizar T. Tchernev

To my parents

ACKNOWLEDGMENTS

The author would like to thank his mentor, Dr. Stephen F. Kingsmore, and the supervisory committee members, Dr. M. Wallace, Dr. E. Wakeland, Dr. E. Sobel and Dr. M. Bubb, for their valuable advice and encouragement.

The author's appreciation is extended to all former and present members of Dr. Kingsmore's laboratory, Dr. Maria D.F.S. Barbosa, Dr. Vishnu S. Mishra, John C. Detter, Dr. Elizabeth B. McMurtrie, Quan A. Nguyen, Sandra M. Holt, Juan M. Teodoro and Andrea Hofig, as well as to Dr. Krishnan Nandabalan and Dr. Madan Kumar from CuraGen Corporation, who contributed to the successful completion of this work.

The author wishes to express his gratitude to Ms. Anna V. Gueorguieva and Mr. Nathan S. Collier for their kindness and continuous support.

Finally, but most importantly, the author would like to thank his parents Lilia Jeleva and Tzvetan Tchernev for their loving care and guidance, his beloved wife Ralitza Gueorguieva for her invaluable support and inspiration, and his fellow-players Ivajlo Kortezov, Rada Kortezova, Ludwig Lubih, Anastas Assenov and Isak Beraha for their true friendship and thought-provoking conversations.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS iv

ABSTRACT vii

INTRODUCTION 1

MATERIALS AND METHODS 22
Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Amplification of CHS Patient and Beige Mouse cDNAs 22
TA Cloning of PCR Products and Plasmid Purification 23
Single-Strand Conformation Polymorphism (SSCP) Analysis 24
DNA Sequencing and Sequence Analysis 24
Allele-Specific Oligonucleotide Analysis 25
PCR Amplification of Mouse *Lyst* cDNA Isoforms 25
Cloning of Human *LYST* cDNAs 26
Molecular Probes 28
Isolation of *LYST2* cDNA Clones 29
Southern Blot Analysis and Autoradiography 30
Northern Blot Analysis 30
Simple Sequence Length Polymorphism (SSLP) PCR Amplification 31
Backcross Mouse Panel and Genetic Mapping 31
Cloning of *LYST* and *LYST2* Bait cDNA Fragments in Yeast Two-Hybrid Vectors 32
Yeast Two-Hybrid Screens 36
Confirmation of the Specificity of Protein Interactions by the Yeast Two-Hybrid System 40

RESULTS 43
Identification of Mutations in Patients with Chediak-Higashi Syndrome and in Beige Mice 43
Identification and Characterization of Lyst mRNA Isoforms 52
Identification and Characterization of *Lyst2*, a Brain-Specific Member of the Chediak-Higashi Syndrome Gene Family 61
Identification and Sequence Analysis 61
Genetic Mapping 62
Expression Analysis 67

Identification of Proteins that Interact with the CHS Protein and with LYST2, Using a Yeast Two-Hybrid Approach 71
IP-1 (EST cg50136.f6, APPENDIX, Sequence 13) 76
IP-2 (EST AA010799, APPENDIX, Sequence 14) 77
IP-3 (EST cg50136.c10, APPENDIX, Sequence 15) 77
IP-4 (EST cg50136.a7, APPENDIX, Sequence16) 78
IP-5 (EST cg50175.c7, APPENDIX, Sequence 17) 79
IP-6 (EST cg50138.g5, APPENDIX, Sequence 18) 80
IP-7 (EST cg50173.d10, APPENDIX, Sequence 19) 81
IP-8 (EST cg50175.h7, APPENDIX, Sequence 20) 82
IP-9 (EST KIAA0192, APPENDIX, Sequence 21) 83
IP-10 (EST cg50136.a5.b, APPENDIX, Sequence 22) 83
IP-11 (EST cg51287.d10, APPENDIX, Sequence 23) 84
IP-12 (EST cg49432.h3.b, APPENDIX, Sequence 24) 85
IP-13 (EST cg50175.c11, APPENDIX, Sequence 25) 86

DISCUSSION 88
Identification and Characterization of *Lyst* mRNA Isoforms 88
Identification of Mutations in Patients with Chediak-Higashi Syndrome and in Beige Mice 92
Identification and Characterization of *Lyst2*, a Brain-Specific Member of the Chediak-Higashi Syndrome Gene Family 94
Identification of Proteins that Interact with the CHS Protein, Using a Yeast Two-Hybrid Approach 98

APPENDIX SEQUENCES OF MOLECULES INTERACTING WITH LYST AND LYST2 123

BIBLIOGRAPHY 177

BIOGRAPHICAL SKETCH 192

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CHARACTERIZATION OF THE CHEDIAK-HIGASHI SYNDROME GENE IN HUMAN AND MOUSE

By

Velizar T. Tchernev

December 1998

Chairman: Stephen F. Kingsmore
Major Department: Pathology, Immunology and Laboratory Medicine

Chediak-Higashi syndrome (CHS) is an autosomal recessive, immune deficiency disorder of human and mouse (beige, *bg*) that is characterized by abnormal intracellular protein transport to and from the lysosome/late endosome. Recent reports have described the positional cloning of homologous genes, *LYST* and *Lyst*, that are mutated in human CHS and in *bg* mice, respectively. However, since the encoded proteins were novel and unlike any of the molecules previously implicated in vesicular trafficking, their identification did not prove immediately helpful in establishing the precise mechanism whereby CHS dysregulates protein transport. Therefore, the main goal of this dissertation was to characterize the Chediak-Higashi gene and its products in more detail using functional genomics approaches. We demonstrated that each of the previously reported *bg* gene sequences are derived from a single gene with alternatively spliced mRNAs. Alternative splicing results in *bg* gene isoforms (*Lyst*-I and *Lyst*-II) that contain different

3' regions. Similarly to mouse, the human mRNA isoforms arise from incomplete splicing and retention of a transcribed intron that encodes the C-terminus of the predicted LYST protein. Additional splicing complexity of smaller isoforms exists. The *Lyst*-III isoform lacks exons α and β, while in *Lyst*-IV, exons α, β and γ are absent. Novel mutations were identified within the coding domain of *LYST* in several CHS patients and *bg* alleles. Interestingly, all *bg* and *LYST* mutations identified to date are predicted to result in either truncated or absent proteins. Mutation and expression analyses suggested that defects in the full-length mRNA alone can elicit Chediak-Higashi syndrome and that expression of the smaller isoform alone cannot compensate for loss of the largest isoform. We have also identified a novel gene in human (*LYST2)* and mouse (*Lyst2*) that appears to be a relative of the *CHS* gene, based on sequence similarity, predicted protein structure and on similar transcript size. Comparison of the relative abundance of *LYST2* and *LYST* mRNAs suggests that *LYST2* is expressed abundantly only in brain and therefore it may represent a brain-specific member of the *CHS* gene family. Using a modified yeast two-hybrid system, a human cDNA library was screened with baits from the coding domains of *LYST* and *LYST2*. Several proteins, which play important roles in protein transport and signal transduction, such as 14-3-3, casein kinase II, calmodulin and Hrs, were found to interact with LYST and/or LYST2. Many of the interacting proteins could be linked in a common pathway that regulates vesicular trafficking and degranulation.

INTRODUCTION

Chediak-Higashi syndrome (CHS) is a primary immunodeficiency, characterized by severe recurrent infections (Blume & Wolff, 1972), partial albinism (Windhorst *et al.*, 1968), nervous system abnormalities (Misra *et al.*, 1991), high incidence of malignancy (Hayakawa *et al.*, 1986) and predisposition to bleeding (Meyers *et al.*, 1974). Eighty-five percent of patients undergo an accelerated lymphoma-like phase, characterized by fever, hepatosplenomegaly, lymphadenopathy, pancytopenia, coagulopathy and widespread lymphohistiocytic organ infiltrates. This complication is fatal in the absence of bone marrow transplantation.

The first 3 cases of Chediak-Higashi syndrome were reported by Beguez-Cesar in 1943, then one case by Steinbrink in 1948, and one case by Higashi in 1954. Chediak published certain aspects of Beguez-Cesar's cases in 1952, and Sato in 1955 recognized the similarities between Chediak's and Higashi's data, inventing the eponym "Chediak-Higashi syndrome." Except for the occasional usage of single names or combinations of the above authors' names, the disease has been most commonly referred to as the Chediak-Higashi syndrome.

Approximately 150 cases of CHS have been reported in the literature. Males and females are affected equally. The mean age of death of CHS patients is 6 years. The Chediak-Higashi syndrome occurs with a low frequency throughout the world. Europe,

America and Japan are the areas from which the majority of the known cases have been described. While most of the reports deal with isolated cases, a unique cluster of Chediak-Higashi syndrome patients has also been published (Ramirez-Duque *et al.*, 1982). Fourteen cases of CHS were found in 12 families living in a defined and relatively isolated geographical area (Pregonero) in the Venezuelan Andes. The patients were pre-school children except one 11-year-old female. Six of the patients were male. All showed the same typical clinical characteristics of the syndrome. Since CHS is an autosomal recessive disease (Sato, 1955), the existence of this high-frequency cluster of CHS patients in a restricted region could be explained by a founder effect in an isolated population with inbreeding.

The presence of giant, perinuclear vesicles with protein markers characteristic of late endosomes and mature lysosomes is pathognomonic for CHS (Burkhardt *et al.*, 1993). The organelles are abnormal both in size (enlarged) and in cytoplasmic distribution (perinuclear). While the late endosomal/lysosomal compartment is uniformly affected in CHS, the histological designation of affected organelles varies by cell type: melanosomes are affected in CHS melanocytes, while secretory granules are affected in CHS neutrophils, eosinophils, mast cells, cytolytic T-cells, proximal renal tubules, and lysosomes are affected in monocytes and fibroblasts. The giant organelles are thought to result from dysregulated fusion of primary endosomes and lysosomes (Burkhardt *et al.*, 1993; Willingham *et al.*, 1981). The morphological abnormalities of *bg* and CHS lysosomes are accompanied by compartmental mis-sorting of proteins such as elastase, β-glucuronidase, and cathepsin G (Holcombe *et al.*, 1994).

The disease also occurs in other mammalian species – mouse, rat, cat, cattle, mink, fox and killer whale. Beige (*bg*) is a well described mouse model of CHS. The clinical and pathological features of CHS patients and *bg* mice are indistinguishable. Prior to the cloning of *bg*, the strongest evidence of homologous defects in beige and CHS came from interspecific genetic complementation studies, which demonstrated that fusion of *bg* and CHS fibroblasts failed to reverse lysosomal abnormalities, while fusion with normal cells did eliminate cytological defects of *bg* or CHS fibroblasts (Perou and Kaplan, 1993; Penner and Prieur, 1987).

Partial albinism is one of the main clinical manifestations of the Chediak-Higashi syndrome. Most patients have hypopigmentation of their hair, skin and ocular fundi, as well as misrouting of the optic nerve fibers that is characteristic of all types of albinism. Windhorst *et al.*, 1966, 1968 reported the presence of giant melanosomes in melanocytes from CHS patients. Some melanosomes were highly irregular, with a suggestion of fusion of separate granules. Most epidermal melanosomes from a CHS patient, observed by Zhao *et al.*, 1994, were several times larger than normal, relatively amelanotic, and occasionally clustered into complexes. No melanosomes were detected in keratinocytes adjacent to melanocytes, suggesting that the normal transport of melanosomes from melanocytes to neighboring keratinocytes is defective. After DOPA histochemistry, used to localize functional tyrosinase, reaction product was observed throughout the cytoplasm, but not in most of the aberrant large melanosomes and complexes, indicating that functional tyrosinase is either not translocated to or is not retained within the giant granules. In contrast to control melanocytes, DOPA-positive vesicles migrated far beyond their usual perinuclear and trans Golgi network location and were seen in the vicinity of

the plasma membrane in melanocytes from CHS patients. When several melanocyte-specific proteins were localized histochemically or immunohistochemically, they were predominantly detected in the perinuclear area and less prevalent in the dendrites of CHS cells, compared to their relatively uniform distribution throughout the cytoplasm of control cells (Zhao *et al.*, 1994), suggesting an inefficient intracellular translocation of these proteins. Medium, conditioned by CHS cells for 1 to 3 days, displayed increasing amounts of tyrosinase activity that did not exist in medium conditioned by numerous control cell lines. β-glucuronidase was also released into the culture medium by CHS melanocytes but not by control melanocytes. These data suggest that the low tyrosinase activity in giant melanosomes results from mislocalization and eventual extracellular secretion of tyrosinase. This hypothesis is supported by the presence of DOPA-positive vesicles beneath the plasma membrane where fusion/secretion could occur.

Patients with Chediak-Higashi syndrome, especially those who survive into adulthood, and elderly beige mice, develop serious neurologic defects. Lockman *et al.*, 1967, reported a patient with the typical clinical and cytological picture of CHS and a severe motor neuropathy. A median nerve biopsy showed that many Schwann cells contained large cytoplasmic particles which the authors considered to be abnormal lysosomes. Two cases with lymphohistiocytic infiltration of the sural and femoral nerves were described by Sung *et al.*, 1969. The myelin sheaths and axons showed swelling and fragmentation or had disappeared. Schwann cells were hypertrophic and contained cytoplasmic inclusions. Ultrastructural examination of dorsal root and sympathetic ganglion cells demonstrated inclusions resembling lysosomes and lipofuscin granules.

Pettit and Berdal, 1984, reported a 25-year-old man with a neurologic disorder that resembled a spinocerebellar degeneration and Parkinsonism. Cerebellar cortical atrophy (Kondo *et al.*, 1994) and diffuse atrophy of the brain with abnormalities in the periventricular and corona radiata regions were detected by computer tomography and magnetic resonance imaging (Ballard *et al.*, 1994). The oldest known CHS patient first seen with a neurologic disorder in early adult life was described by Uyama *et al.*, 1994. This 39-year-old woman developed mental deterioration, Parkinsonism, muscular atrophy of limbs, and loss of tendon reflexes. MRI showed marked temporal dominant brain atrophy and diffuse spinal cord atrophy. Similarly, beige mice that survive to 17 months of age show a progressive neurologic disorder accompanied by nearly complete loss of cerebellar Purkinje cells (Murphy and Roths, 1978). Guo *et al.*, 1992, reported cytoarchitectonic abnormalities in the brains of beige mice. In the cerebellum, Purkinje cells and clusters of granule cells, and occasionally glia cells, were located ectopically in the molecular level. In the hippocampus, ectopically situated pyramidal cells were found. The infrapyramidal mossy fiber layer was not compact, but appeared as clumps of scattered fiber bundles. Pezeshkpour *et al.*, 1986, reported the presence of giant lysosomes in Schwann cells in a nerve biopsy from a CHS patient. Peripheral neuropathy in a 33-year-old brother and a 29-year-old sister with CHS were described by Misra *et al.*, 1991. Both showed evidence of a sensory neuropathy associated with central nervous system involvement. Nerve conduction studies indicated an axonal neuropathy. Sural nerve biopsy demonstrated a loss of myelinated nerve fibers, particularly those of larger size, and of unmyelilnated axons. By light and electron microscopy, granulated cells with the staining properties of mast cells were present in the endoneurium. Instead of being the

usual 0.1-0.2 μm in diameter, the granules were approximately 10 times larger and with unusual appearance. Blume and Wolf, 1972, reported irregular giant inclusions, thought to be lysosomes, in the cytoplasm of Schwann cells. These morphologic observations of abnormally large granules suggest that the neurological aberrations in CHS result from defective vesicular transport of molecules that are important for the normal cellular function.

Bleeding tendency is another major clinical manifestation in patients and animals with CHS. While the platelet numbers are normal (Novak *et al.*, 1985; Phillips *et al.*, 1967), dense platelet granules are virtually absent, whole-blood serotonin is markedly decreased, and platelets have reduced ability to accumulate serotonin (Meyers *et al.*, 1974; Meyers *et al.*, 1983). The prolonged bleeding time and low serotonin concentrations were converted to normal values following transplantation of normal bone marrow in beige mice. Likewise, control mice displayed symptoms of storage pool deficiency when transplanted with mutant marrow (Novak *et al.*, 1985). These studies demonstrate that the CHS platelet storage pool deficiency results from a defect in bone marrow precursor cells. In agreement with this conclusion, Menard and Meyers, 1988, showed that the dense granule precursors in maturing and mature megakaryocytes were absent. Abnormal platelet aggregation was described in 6 patents with CHS (Apitz-Castro *et al.*, 1985). The total content and the maximal amounts of the dense granule constituents secreted after thrombin stimulation were greatly decreased. CHS platelets loaded with radiolabelled serotonin showed a spontaneous release of radioactivity not observed in normal platelets under the same conditions. Similar defects were reported in cats (Colgan

et al., 1989) and in blue foxes (Sjaastad *et al.*, 1990) with Chediak-Higashi syndrome. In beige mice, platelet aggregation after stimulation was significantly decreased compared to that of normal mice (Pratt *et al.*, 1991). Platelets from beige mice were approximately 10 times more sensitive to prostacyclin inhibition of collagen-induced aggregation than control platelets. The stores of serotonin and adenine nucleotides were decreased. Rendu *et al.*, 1983, reported a large reduction in the number of serotonin-storage granules (dense bodies) but otherwise normal ultrastructure and normal amounts of α- and catalase-positive granules in 3 patients with Chediak-Higashi syndrome. The platelet release reaction, studied with thrombin as an inducer, was impaired. The serotonin uptake by the patients' platelets was low and its metabolism was increased. These findings show that human CHS platelets are deficient in the storage pool of dense granule substances and suggest that this granule defect influences the release mechanism of other granule constituents, leading to abnormal aggregation and bleeding predisposition.

Most morbidity and mortality in CHS result from immune deficiency, with severe defects in innate immunity (granulocytes and NK cells) and relatively mild impairment of acquired immune function (B and T cells).

The initial immune defect recognized in CHS was a striking reduction in the ability of granulocytes to kill phagocytosed microorganisms (Blume *et al.*, 1968; Gallin *et al.*, 1974). Phagocytosis of radiolabelled *Staphylococcus aureus* by CHS granulocytes was normal. However, there was reduced bactericidal activity of *S. aureus* and group D *Streptococcus* by CHS cells through 90 min of incubation. This defect was most pronounced at early time periods and was related to impaired intracellular killing. There

are occasional reports suggesting more widespread neutrophil abnormalities. In one study (Bellinati-Pires *et al.*, 1992), in which phagocytosis and killing of *Staphylococcus aureus* were simultaneously evaluated by a fluorochrome phagocytosis assay, both appeared to be deficient in neutrophils form two patients with CHS. The presence of giant phagolysosomes, enclosing bacteria in active proliferation 45 min after phagocytosis was noteworthy, and corresponded with the impaired bactericidal activity of these leukocytes. In an under-agarose assay, phagocytes of homozygote and phenotypically normal heterozygote CHS cats recognized and responded equally well to bacterial stimuli as did cells from control animals but traveled shorter distances primarily because of a reduced inherent motility. Similar results were obtained when phagocyte chemotaxis was evaluated with zymosan-activated serum (Colgan *et al.*, 1992). Mildly defective CHS neutrophil and monocyte locomotion was discovered by the micropore filter technique using modified Boyden chambers (Yegin *et al.*, 1983). Isolated reports suggest that diminished surface expression of C3bi receptor and decreased oxygen radical generation by polymorphonuclear leukocytes (PMNs) might be some of the factors to explain the moderately defective PMN mobility, chemotaxis and bactericidal activity (Cairo *et al.*, 1988; Kubo *et al.*, 1987). While the literature contains some conflicting data, the consensus is that, with the exception of the ability to kill phagocytosed microorganisms, other granulocyte activities, such as phagocytosis, chemotaxis and protein degradation, are relatively well preserved in CHS (Gallin *et al.*, 1974; Clawson *et al.*, 1978). Analogous defects occur in CHS monocytes and macrophages -- cytolysis of tumor targets is impaired, while phagocytosis and chemotaxis are almost normal (Gallin *et al.*, 1975; Mahoney *et al.*, 1980). CHS selectively involves crystalloid granules in late

eosinophils, while the non-crystalloid granules are preserved, suggesting that the underlying pathology does not affect all lysosomal subpopulations (Hamanaka *et al.*, 1993).

The second category of CHS immune defects is a virtual ablation of natural killer (NK) cell activity and antibody dependent cellular cytotoxicity (Roder and Duwe, 1979; Roder *et al.*, 1979, 1982). Mature NK cells demonstrate low lytic efficiency, assessed by conventional 4 h chromium release assay and by a more recently developed single cell liquid cytotoxic assay. The NK cells were not activated by prolonged *in vitro* incubation with IFN-γ, but they exhibited IL-2 induced cytotoxicity, though the magnitude of induction was uniformly less than in controls (Virelizier and Griscelli, 1980; Holcombe, 1992). CHS NK cells have the necessary cellular structures required for their role as lytic effector cells, but lack cytotoxic function due to a relative refractoriness in initiating the post-binding lytic mechanism. NK cells in CHS patients are blocked at a post-recognition, post-activation step in the cytolytic pathway, subsequent to the burst of oxygen intermediates but preceding the lethal hit. In summary, while NK cells are present in normal numbers, bind target cells effectively and generate superoxide ions normally, exocytosis of lytic granules from the CHS NK cells is refractory (Targan and Oseas, 1983; Roder *et al.*, 1983; Brahmi, 1983).

Tumor susceptibility of *bg* mice was examined as a direct test of the hypothesis that NK cells are involved in surveillance against neoplasia. Subcutaneous inoculation in a transplantation test yielded higher tumor take incidence in *bg/bg* mice than in heterozygous littermates. The differences were most striking in the early phase of the

observation period, with a majority of the total tumor takes in *bg/bg* registered within 2 weeks, whereas only a small part of the tumors in *bg/+* mice were palpable at this time. Thus, in addition to the increased incidence in homozygous mice, the progressively growing tumors appeared with a shorter latency than in controls. This was also reflected in the larger mean tumor diameter by 2 and 3 weeks and the earlier deaths among tumor-bearing *bg/bg* than *bg/+* mice. These findings correlated well with the *in vitro* measured splenic NK activity against both tumors tested. Tumor cells had a low, but significant, sensitivity to *bg/+* spleen cell cytotoxicity (5-10 % specific lysis), whereas *bg/bg* spleen consistently gave values below 3% specific lysis (Kaerre *et al.*, 1980). In another study, a tumor line modified to be sensitive to NK cytotoxicity *in vitro*, demonstrated *in vivo* an increased growth rate, faster induction time and an enhanced metastatic capability in *bg* compared to control mice. This was not found with a tumor line insensitive to NK activity. *In vivo* activation of NK cells in *bg* and control mice resulted in decrease in tumor growth rate and metastatic frequency (Talmage *et al.*, 1980). In a large Japanese cohort, non-Hodgkin's lymphoma occurred in over one third of CHS patients (Hayakawa *et al.*, 1986). These results indicate that NK cells have an important function in the host's control of tumor growth and metastasis. The idea to use the *bg* model to dissect the contribution of NK cells to immunity, and in particular, to tumor immunosurveillance was dispelled somewhat, however, by the subsequent recognition of concomitant impairment, albeit mild, in T and B cell function.

Although both B and T cells are affected in CHS, the abnormalities in the specific defense mechanisms are not as severe as those in natural immunity and have less importance for the pathogenesis of CHS. The pathognomonic giant cytoplasmic vesicles

are detected in B cells from CHS patients following activation and differentiation (Grossi *et al.*, 1985). Moderate and variable reduction of LPS-induced B cell mitogenesis is demonstrated in CHS. The *in vivo* anti-trinitrophenyl antibody response to a thymus-independent antigen (TNP-Ficoll) was found to be significantly lower in C57BL/6J-*bg* than in C57BL/6J mice, although both strains responded similarly to an analogous thymus-dependent antigen (TNP-ovalbumin) (Pflumio *et al.*, 1990). Since the marginal zone macrophages of the spleen were previously shown to be essential for the initiation of antibody responses to thymus-independent antigens, they might be another target of the *bg* mutation, possibly resulting in impaired antigen presentation to B cells.

It is notable that homozygosity at the *bg* locus confers protection against immune complex glomerulonephritis to BXSB mice with systemic autoimmune disease (Clark *et al.*, 1982). However, this phenomenon is probably not due to a decreased B cell function because the levels and types of autoantibodies remain unchanged in *bg* homozygotes.

Extremely high antibody titers to the Epstein-Barr virus specific viral capsid antigen, to the restricted component of the EBV-induced early antigen complex and to the EBV-associated nuclear antigen were detected in CHS patients (Merino *et al.*, 1983). It is likely that the elevated antibody titers reflect an increased production of these antigens due to defective NK and antibody-dependent cellular cytotoxicity (ADCC) activities.

Cytotoxic lymphocytes (CTLs) from CHS patients are unable to destroy target cells recognized via the T cell receptor (TCR). Individual CTL clones show poor killing that can be increased in longer assays. However, in the presence of cycloheximide, the small amount of killing is abolished, indicating that killing arises from newly synthesized proteins, rather than from proteins stored in granules. It has been shown that the CHS

CTL clones express normal levels of the lytic proteins granzyme A, granzyme B and perforin, which are processed properly during biosynthesis and targeted correctly to giant lytic granules. Despite the difference in size, CHS and normal lytic granules are similar, both contain the lysosomal enzyme cathepsin D and granzyme A, and lack the mannose-6-phosphate receptor. However, unlike normal CTL clones, the CHS CTLs are unable to secrete their giant granules in which the lytic proteins are stored. After cross-linking the TCR, CHS CTL clones fail to secrete granzyme A, as assayed by both enzyme release and confocal microscopy (Baetz *et al.*, 1995).

The proportion of gamma-delta T cells among peripheral blood mononuclear cells is significantly increased in CHS. The cellular machinery for lysis of target cells *in vitro* is present in CHS-derived gamma-delta T cell clones (Holcombe *et al.*, 1990). This is the first example of a specific immunodeficiency disorder with a relative expansion of these T cells.

A short summary of the immunological features of CHS is presented in Table 1.

Table 1. Main immunologic features in Chediak-Higashi syndrome

INNATE IMMUNITY		ACQUIRED IMMUNITY	
Granulocytes	NK cells	B cells	T cells
1. Greatly reduced ability to kill phagocytosed micro-organisms 2. Chemotaxis, phagocytosis and other activities are relatively well preserved	1. Very low lytic activity caused by defective exocytosis of lytic granules 2. Numbers of NK cells, binding to target cells, and oxygen radical production are normal	1. Reduced antibody response to thymus-independent antigens (may be due to impaired natural immunity) 2. Mildly reduced proliferative response to LPS	1. Decreased ability of CTLs to kill target cells, caused by defective secretion of lytic granules 2. Relative expansion of gamma-delta T cell population
PATHOGNOMONIC GIANT PERINUCLEAR VESICLES			

Most mortality in CHS patients is caused by two manifestations of their immune deficiency -- severe pyogenic infections and the accelerated phase. Lung, upper respiratory tract and skin infections caused by *Staphylococcus, Streptococcus* and *Pneumococcus* predominate in CHS, although a wide variety of infection sites and organisms occur. The accelerated phase is considered a clinicopathologic expression of a virus-associated lymphoproliferation because of low $CD4^+$ to $CD8^+$ T lymphocyte ratio, presence of Epstein-Barr virus genome in the mononuclear cells of the lymph node, blood and bone marrow, and possible clinical responses to acyclovir. The current treatment for CHS patients ranges from vitamins to cytostatic agents and bone marrow transplantation, but the prognosis remains poor because of increasing susceptibility to infections and progressive neurologic deterioration. Several studies suggest novel strategies for immunomodulation: recombinant canine granulocyte colony-stimulating factor to improve the neutrophil function (Colgan *et al*., 1992), IL-2 to induce cytotoxicity of CHS lymphocytes (Holcombe, 1992), and high-dose intravenous gammaglobulin for management of the accelerated phase (Kinugawa and Ohtani, 1985). The goal of such therapies would be to partially compensate the immune cell dysfunction caused by the basic genetic defect and manifested as abnormal protein trafficking.

The transport of newly synthesized proteins from the endoplasmic reticulum to the Golgi apparatus and from there to the cell surface and elsewhere, is mediated by transport vesicles, which transfer enclosed proteins by cycles of vesicle budding and fusion. The abnormalities in CHS can be explained by a compartmental mis-sorting of a variety of proteins, which are necessary for the proper function of the affected cell types, for example CHS CTLs cannot secrete stored granzyme A and other lytic proteins, and

therefore cannot lyse target cells efficiently. The morphologic manifestation of the protein sorting defect in CHS is the presence of giant perinuclear vesicles in the cytoplasm of the affected cells. The assortment of protein trafficking defects observed in CHS defies simple explanation: trafficking both to and from the late endosomal/lysosomal compartments is affected, and proteins which utilize both the cation-independent mannose-6-phosphate and other transport mechanisms are mis-sorted. In contrast, targeting of membrane proteins to lysosomal/late endosomal compartments (LAMP1, LAMP2, rab7, acid phosphatase) is intact. Experiments demonstrating spontaneous concavalin-A induced capping of CHS granulocytes suggest that the defect impairs interactions between microtubules and endosomes/lysosomes (Oliver *et al.*, 1975). However, detailed examination of microtubules and microtubule-based motors has not disclosed an abnormality (Perou and Kaplan, 1993). In summary, abnormalities in lysosomal and endosomal processing may explain the clinical features of CHS, suggesting that the defect involves a protein that regulates certain aspects of trafficking to and from lysosomal/late endosomal compartments in many cell types.

A summary of the mechanisms involved in the pathogenesis of Chediak-Higashi syndrome is presented in Fig.1. All clinical manifestations are caused by a common defect in the regulation of vesicular trafficking in different cell types: melanosomes are affected in melanocytes, resulting in partial albinism; dense granules in platelets are defective, leading to bleeding predisposition; lysosomes and secretory vesicles in numerous immune cells are deficient, causing frequent infections and increased incidence of tumors; and giant inclusions are present in Schwann and Purkinje cells, which could be associated with neurological abnormalities.

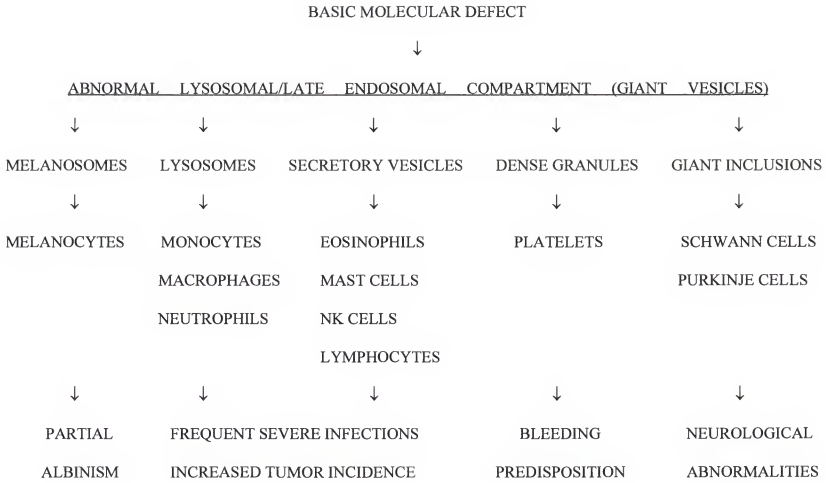

Figure 1. Mechanisms involved in the pathogenesis of Chediak-Higashi syndrome.

The need for molecular understanding of the basic defect in Chediak-Higashi syndrome led to the first mapping studies, which were greatly facilitated by the existence of the mouse model beige. Linkage of satin (*sa*) with *bg* defined a new linkage group, later found to lie on proximal mouse chromosome (Chr) 13 (St. Amand and Cupp, 1958; Lyon and Meredith, 1969; Lane, 1971). Owen *et al*., 1986, showed that in mouse, *bg* is closely linked to the gene for the γ-chain of the T cell receptor (TCR). However, in 3 families with CHS, RFLPs in the TCR γ-gene were inherited discordantly with CHS, demonstrating non-linkage (Holcombe *et al*., 1987) and suggesting a break in the region of homology between human chromosome 7p and mouse chromosome 13. Further refinement of the genetic map of proximal mouse Chr 13 (Justice *et al*., 1990) led to linkage of *bg* to the gene encoding nidogen (*Nid*), a structural basement membrane protein (Jenkins *et al*., 1991). No crossovers between *Nid* and *bg* were detected in 123 mice, suggesting that the two loci are within approximately 2.4 cM. Since human nidogen (*NID*) has been assigned to Chr 1q43 (Olsen *et al*., 1989), the close linkage between *bg* and *Nid* established a new group of homology between mouse Chr 13 and human Chr 1q, suggesting that the gene responsible for the Chediak-Higashi syndrome is located in the telomeric region of human chromosome 1q. These results confirmed the earlier observation of non-linkage between CHS and TCR-γ (Holcombe *et al*., 1987) and the existence of a breaking point in the region of homology between mouse Chr 13 and human Chr 7p. Indeed, linkage between *NID* and CHS on human chromosome 1q was reported by Goodrich and Holcombe in 1995. Homozygosity mapping using markers derived from distal human chromosome 1q was carried out by Fukai *et al*., 1996, in four

inbred families or probands with typical childhood Chediak-Higashi syndrome. The human *CHS* gene was localized to an 18.8 cM interval, flanked by D1S446 and D1S184, on Chr 1q42-q44. Barrat *et al.*, 1996, studied 10 CHS patients from 9 families and mapped the CHS locus to an approximately 5 cM region between D1S163 and D1S2680 on human chromosome 1q42.1-q42.2. As a first step in a positional cloning effort, Kingsmore *et al.*, 1996a, precisely localized the beige gene to a 0.24 cM interval on proximal mouse Chr 13 by segregation analysis. In a total of 726 mice from 3 different backcrosses, *bg* was found to cosegregate with 6 genetic markers (*Nid, Estm9, D13Mit56, D13Mit162, D13Mit237* and *D13Mit240*). The candidacy of *Nid* and *Estm9* was evaluated by Southern, northern and RT-PCR analyses, which suggested that neither of these two genes represented the *bg* gene. However, interstrain differences in pulsed field restriction fragment lengths were observed, providing a useful and simple test for initial evaluation of candidate genes. A refined physical map of the *bg* critical region was reported by Kingsmore *et al.*, 1996b. Interspersed repetitive element-polymerase chain reaction (IRE-PCR) and direct cDNA selection were used to identify 20 novel sequence tag sites which enabled the construction of 2 contigs composed of overlapping YAC and P1 clones, covering approximately 2400 kb of the *bg* non-recombinant interval. In 1996, two groups simultaneously reported the identification of the homologous beige and Chediak-Higashi syndrome genes. Barbosa *et al.*, 1996, isolated a gene designated *Lyst* (*Lys*osomal *t*rafficking regulator), from YAC 195A8 by direct selection with mouse spleen cDNA. It was shown that *Lyst* represented the beige gene with a complete open reading frame of 4635 nucleotides, encoding a protein of 1545 amino acids. It was also demonstrated that corresponding human clones with high degree of homology to mouse

Lyst represented the gene responsible for Chediak-Higashi syndrome. Perou *et al.*, 1996b, also reported the identification of the beige gene by YAC complementation and positional cloning. Interestingly, the partial cDNA sequence isolated by Perou *et al.*, did not match or even overlap with the sequence of *Lyst*, described by Barbosa *et al.* One of the aims of the present dissertation was to explain this discrepancy. This issue was resolved by the demonstration that each of the previously reported beige gene sequences was derived from a single gene by alternative splicing (Barbosa *et al.*, 1997; Nagle *et al.*, 1996), as described in the RESULTS section of this dissertation. However, the expectations that the identification of the complete mouse and human cDNA sequences of the CHS gene would elucidate the precise function of the encoded protein, were not met. The 11403-bp full-length open reading frame corresponded to a polypeptide of 3,801 amino acids, with an estimated molecular mass of 429,153 daltons. Although several putative domains were found in Lyst, the predicted protein was novel and unlike any of the molecules previously implicated in vesicular transport (Barbosa *et al.*, 1996, 1997; Perou *et al.*, 1996b; Nagle *et al.*, 1996). A helical region of Lyst was similar to stathmin, a phosphoprotein that regulates the polymerization of microtubules and acts as a relay for intracellular signal transduction. Therefore, this region of *Lyst* potentially encodes a protein interaction domain that may regulate microtubule-mediated lysosome trafficking. Although the CHS protein (LYST) is not predicted to have transmembrane helices, the carboxy-terminal tetrapeptide (CYSP) of the mouse Lyst-II isoform is similar to known prenylation sites and may provide attachment to membranes through thioester linkage with the cysteine. A novel conserved domain, designated BEACH (BEige And CHs), was defined based on homology with open reading frames from *S. cerevisiae, C. elegans* and

a human cell division control protein 4-related protein (CDC4L). The function of the BEACH domain remains unknown. Helical regions in the CHS protein resemble ARM and HEAT repeat motifs, which tend to form long rods. Some HEAT-containing proteins are associated with vesicular trafficking but the functional significance of these motifs in Lyst is unclear. The C-terminal region of the CHS protein contains several consecutive WD40 motifs, which were originally identified in the β-subunit of the G-protein transducin. This WD40 region may form β-sheets arranged in a propeller-like structure that is thought to mediate protein-protein interactions. Based on the presence of ARM, HEAT, BEACH and WD40 domains, it has been proposed that, similarly to the yeast serine/threonine protein kinase VPS15, Lyst functions as a component of a membrane-associated signal transduction complex that regulates intracellular protein trafficking. The CHS protein contains multiple sites of potential phosphorylation by casein kinase II (CKII), PKC, cAMP-dependent protein kinase and tyrosine kinase. Lyst seems to contain helical bundles with clusters of phosphorylation sites at their ends. Phosphorylation of these positions could provide a control mechanism by causing conformational change in the bundles, thereby affecting interactions with other molecules. In summary, the domains identified in Lyst support its potential involvement in the regulation of signal transduction and protein trafficking but do not place the CHS protein in any known family and do not define any proteins that interact with Lyst, leaving the question about its mechanism of action unanswered.

Therefore, the main objective of this dissertation was to characterize the Chediak-Higashi syndrome gene and its products in more detail, and to identify proteins that

interact with Lyst, in order to get some insights about its biological role. Following the cloning of the *CHS* gene in human and mouse, the following specific aims for the present dissertation were determined:

1. Identification of mutations in patients with Chediak-Higashi syndrome and in beige mice.
2. Identification and characterization of *Lyst* mRNA isoforms.
3. Identification and characterization of *Lyst2*, a brain-specific member of the Chediak-Higashi syndrome gene family.
4. Identification of proteins that interact with the CHS protein and with LYST2, using a yeast two-hybrid approach.

Detection of mutations in the *CHS* gene would help the identification of critical amino acids and domains in Lyst and would enable genotype-phenotype correlation, as well as the design of tests for early diagnosis of Chediak-Higashi syndrome. The characterization of different mRNA isoforms would determine the functional importance of individual alternatively spliced variants and their expression in different tissues. The identification of proteins that interact with Lyst would reveal the pathways in which the CHS protein is involved and would enhance our knowledge about the mechanisms of regulation of protein trafficking, particularly to and from the lysosomes.

MATERIALS AND METHODS

Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Amplification of CHS Patient and Beige Mouse cDNAs

Lymphoblastoid and fibroblast cell lines from CHS patients were obtained from the Coriell Institute for Medical Research (Camden, NJ). Mice carrying different alleles of the beige mutation and control mice were obtained from The Jackson Laboratory (Bar Harbor, ME), sacrificed, and their organs were harvested and homogenized. Total cellular RNA was isolated from human and mouse cells by extraction with phenol/guanidine isothiocyanate (TRIzol Reagent, Life Technologies Inc., Gaithersburg, MD). One to three μg of RNA were reverse transcribed using SuperScript II RNase H^- reverse transcriptase (Life Technologies Inc.), with oligo(dT)-primed first strand cDNA synthesis. DNA fragments of interest were PCR amplified with KlenTaq DNA polymerase (Clontech Laboratories, Palo Alto, CA) using the following amplification conditions: 94°C for 2 min, then 30 cycles of 94°C for 20 s, 55°C for 15 s, 68°C for 1 min, followed by 6 min at 68°C. PCR products were separated by electrophoresis on agarose gels, visualized by ethidium bromide staining and purified using Wizard PCR Preps DNA Purification System (Promega, Madison, WI). One μL of purified PCR product was used as template for a hemi-nested PCR amplification with KlenTaq DNA polymerase under the same

conditions as above. PCR products were separated by electrophoresis on agarose gels, visualized by ethidium bromide staining and purified using Wizard PCR Preps DNA Purification System. The DNA concentration in each sample was determined using a Spectronic 1001 spectrophotometer (Milton Roy/Bausch & Lomb, Rochester, NY) and the purified PCR products were sequenced either directly or following TA cloning and plasmid purification.

TA Cloning of PCR Products and Plasmid Purification

PCR products were cloned in the pCR 2.1 vector using the TA Cloning kit (Invitrogen, San Diego, CA) according to the manufacturer's instructions. Briefly, PCR products were ligated in pCR 2.1 vector overnight at 14°C and INVαF' competent *E. coli* cells were transformed with the constructs. Cells were grown overnight at 37°C on LB plates containing 50 μg/mL kanamycin and 40 mg/mL X-Gal, then liquid LB/kanamycin cultures were inoculated with positive colonies and incubated overnight with shaking (RPM=225) at 37°C, and plasmids were isolated using the Wizard Plus Minipreps DNA Purification System (Promega). Purified plasmids were visualized on ethidium bromide-stained agarose gels, the DNA concentration in each sample was determined using a Spectronic 1001 spectrophotometer and ~0.3 μg were used for sequencing.

Single-Strand Conformation Polymorphism (SSCP) Analysis

Detection of nucleotide changes by SSCP was performed as described by Aberanthy *et al.*, 1997. Briefly, each PCR product was mixed with an equal volume of denaturing buffer and heated to 95°C for 3 min. The samples were loaded onto 0.8 mm thick, 10% native polyacrylamide gels which were run at ambient temperature at 9 W for 6-10 h, depending on the size of the PCR product. Bands were visualized by silver staining (Beidler *et al.*, 1982).

DNA Sequencing and Sequence Analysis

An ABI PRISM 310 Genetic Analyzer and dye terminator cycle sequencing kit (Perkin Elmer, Foster City, CA) with AmpliTaq DNA polymerase, FS were used for sequencing according to the manufacturer's instructions. Both strands of purified PCR products and plasmids were sequenced to confirm the accuracy of the results. Sequences were analyzed with the GCG Package (Devereux *et al.*, 1984) and searches of the National Center for Biotechnology Information database were performed using the BLAST network server (Altschul *et al.*, 1990) (National Library of Medicine, via INTERNET) and the Whitehead Institute Sequence Analysis Program (MIT, Cambridge, MA). For analysis of sequences, identified by the yeast two-hybrid method as potential interacting proteins, see the "Yeast Two-Hybrid Screens" section later in this chapter.

Allele-Specific Oligonucleotide Analysis

PCR products spanning the mutation site in patient 371 were transferred to nylon membranes using a slot-blot apparatus. Approximately 5 ng of each PCR product was treated with a denaturing solution (0.5 M NaOH, 1.5 M NaCl), split in half and loaded in duplicate. Two 17mer oligonucleotides that span the region containing the mutation were synthesized. One contained the sequence of the normal allele (5'-CGCACATG GCAACCCTT-3'), while the other contained the sequence of the mutant allele (5'-GCACATGGGCAACCCTT-3'). These were end-labeled with [γ-^{32}P]dATP using T4 polynucleotide kinase and hybridized to the membranes at 50°C. Hybridization and wash buffers were as described (Church and Gilbert, 1984). Membranes were washed sequentially at 45 °C, 55 °C and 65°C for 10 min each and exposed to X-ray film.

PCR Amplification of Mouse *Lyst* cDNA Isoforms

Total cellular RNA was isolated from mouse bone marrow cells by extraction with phenol/guanidine isothiocyanate (TRIzol Reagent). Five µg of RNA were reverse transcribed using SuperScript II RNase H$^-$ reverse transcriptase, with first strand cDNA synthesis priming by an oligo(dT) primer and a gene-specific primer (5'-CTCGCAGAG CGGTGCTTATGTCCTGTG-3', 5'-CACAGTCATGGGACTGCTAA-3', for *Lyst*-I and *Lyst*-II, respectively). One µL of cDNA was used as template for PCR amplification using KlenTaq DNA polymerase. PCR conditions were: 94°C for 1 min, then 30 cycles of 94°C for 15 s, 68 °C for 10 or 5 min (for *Lyst*-I and *Lyst*-II, respectively), followed by 20

or 10 min (for *Lyst*-I and *Lyst*-II, respectively) at 68 °C. The same upstream primer (5'-AGCGGAGGTGAAGCCTTATGCTGAGACAGT-3') was used for PCR amplification of both isoforms. Isoform-specific downstream primers used were 5'-TTATGTCCTGTGGGGACACTCCTTC-3' (for *Lyst*-I) and 5'-ACAGAGCATCC CCACTTCCCTATCTAAAGT-3' (for *Lyst*-II). The 2 shorter mouse isoforms (*Lyst*-III and *Lyst*-IV) were PCR amplified from mouse melanocyte cDNA using KlenTaq DNA polymerase. PCR conditions were: 94°C for 2 min, then 36 cycles of 94°C for 20 s, 56°C for 15 s, 68°C for 2.5 min, followed by 10 min at 68°C. PCR products were separated by electrophoresis on agarose gels and visualized by ethidium bromide staining. One μL of 1:500 dilution of PCR products was used as template for a nested PCR amplification with KlenTaq DNA polymerase. PCR conditions were: 94°C for 2 min, then 33 cycles of 94°C for 20 s, 56°C for 6 s, 68°C for 3 min, followed by 6 min at 68°C. PCR products were separated by electrophoresis on 1.2% low melting point agarose gels, visualized by ethidium bromide staining, excised from the gel with a sterile blade and purified using Wizard PCR Preps DNA Purification System. Purified PCR products were cloned in pCR 2.1 vector by the TA method, and sequenced.

Cloning of Human *LYST* cDNAs

Normal human bone marrow poly(A)$^+$ RNA was obtained from Clontech Laboratories and 1 μg was reverse transcribed using SuperScript II RNase H$^-$ reverse transcriptase, with a *LYST*-specific primer (5'-CTTGTTGGCTAGTGCATATTGA CACAATCTTCC-3') used for first strand cDNA synthesis. The resulting *LYST*-specific

cDNA was used as a template for PCR amplification of the *LYST* coding domain with KlenTaq DNA polymerase in 50 μL reactions. PCR conditions were: 94°C for 2 min, then 33 cycles of 94 °C for 20 s, 63 °C for 15 s (for some reactions) and 68°C for 10 min, followed by 12 min at 68°C. PCR products were separated on ethidium bromide-stained low melting point 1% agarose gels, excised with a sterile blade, purified using Wizard PCR Preps Purification System, and cloned in pCR 2.1 vector by the TA method. Some PCR products were treated with 10 U exonuclease I (United States Biochemicals) for 15 min at 37 °C and reamplified by a nested PCR with KlenTaq DNA polymerase prior to cloning in the pCR 2.1 vector. The identity of the cloned products was confirmed by PCR amplification from purified plasmids with *LYST*-specific primers using Taq DNA polymerase in 20-μL reactions. PCR conditions were: 94°C for 2 min, then 30 cycles of 94°C for 20 s, 56 °C for 10 s, 72°C for 1 min, followed by 5 min at 72 °C. PCR products were visualized on ethidium bromide stained agarose gels and photographed.

Segments of the human *LYST* cDNA sequence were obtained by an anchored, nested PCR (5'RACE-PCR) using liver cDNA (Clontech Laboratories) as a template, by RT-PCR using total RNA and by sequencing of human expressed sequence tags (EST) similar in sequence to mouse *Lyst*. For the 5'RACE-PCR, 2 nested primers were derived from a human EST (GenBank accession No. W26957) and had the following nucleotide sequences: 5'-CCAAGATGAAAGCAGCCGATGGGGAAAACT-3' and 5'-TCAGCCTCTTTCTT GCTCCGTGAAACTGCT-3'. For RT-PCR experiments, total RNA was prepared from the promyelocytic HL-60 cell line. RT-PCR was performed with Expand polymerase (Boehringer Mannheim) with the following primer pairs: 5'-

AGTTTATGAGTCCAAATGAT-3' and 5'-GAATGATGAAGTTGCTCTGA-3' (bp 490-2034); 5'-CAGCAGTTCTTCAGATGGA-3' and 5'-ATCTTTCTGTTGTTCCC CTA-3' (bp 1891-3050); and 5'-TAGGGGAGCAACAGAAAGAT-3' and 5'-GCTCAT AGTAGTATCACTTT-3' (bp 3320-4722). The primers used to amplify the cDNA between base pairs 1891 and 3050 were derived from the mouse *Lyst* sequence. Human primers were designed from the sequence of the PCR product (1159 bp) and used to amplify the flanking sequences. Human *LYST* intron σ' was PCR amplified from human genomic DNA (100 ng) with the primers 5'-CCGCTCGAGTAGGATCTT TAAGGTGAATAAC-3' and 5'-GTGATACTACTATGAGCCCTTCACAGTATC-3'. PCR conditions were: 94°C for 1 min, then 32 cycles of 94 °C for 15 s, 63 °C for 20 s and 72 °C for 30 s, followed by 10 min at 72 °C.

Molecular Probes

The *LYST2* probe was PCR amplified from a human fetal brain partial length cDNA clone (#32273), obtained from Image Consortium, using internal primers. KlenTaq DNA polymerase was used for amplification and the PCR conditions were as follows: 94°C for 2 min, then 36 cycles of 94 °C for 30 s, 54 °C for 15 s, 68 °C for 5 min, followed by 5 min at 68 °C. PCR product was then purified with Wizard PCR Preps DNA Purification kit. Other probes are described in the text or in the legends of the corresponding figures. All probes were radio-labeled by the hexanucleotide technique (Prime-It II random primer kit, Stratagene) with [α-^{32}P]dCTP (Amersham) according to the manufacturer's instructions.

Isolation of *LYST2* cDNA Clones

LYST2 clones were isolated by screening cDNA libraries using vector-insert PCR and hybridization. A human fetal brain cDNA library (Life Technologies Inc.) was screened using hemi-nested PCR with one vector primer and one *LYST2*-specific primer. One μL of cDNA library ($5x10^9$ cfu/ml) was used as template for PCR amplification with KlenTaq DNA polymerase. PCR conditions were: 94°C for 5 min, then 33 cycles of 94°C for 30 s, 63 °C for 30 s, 68 °C for 3.5 min, followed by 10 min at 68 °C. PCR products were separated by electrophoresis on agarose gels and visualized by ethidium bromide staining. PCR products were excised from low melting agarose gels, purified (Wizard PCR Preps DNA Purification kit), and cloned using the TA method. Resulting plasmids were purified with Wizard Plus Minipreps DNA Purification kit and sequenced.

Membranes containing clones from a human fetal brain cDNA library (Clontech Laboratories) were screened by hybridization with a radioactively labeled cDNA probe from the 3'-end of the *LYST2* coding domain. Positive clones were detected by phosphorimaging (Molecular Dynamics), and purified plasmids were isolated and sequenced. Then new probes were PCR amplified from the 5'-end of the novel cDNA sequence and the library screening was repeated in order to isolate overlapping clones extending the sequence in the 5' direction.

A ZAP Express *Eco*RI mouse embryo cDNA library was screened by hybridization with the *LYST2* probe. Identified cDNA clones were purified using Wizard Plus Minipreps DNA Purification kit and sequenced.

Southern Blot Analysis and Autoradiography

Genomic DNA was isolated from mouse organs using standard techniques (Sambrook *et al.*, 1989) or obtained from The Jackson Laboratory (Bar Harbor, ME). DNA was digested with restriction endonucleases from Promega (Madison, WI) or Boehringer Mannheim (Indianapolis, IN), and 10-μg samples were subjected to electrophoresis in 0.85% agarose gels for ~ 20 h at 35V. DNA was transferred to ZetaProbe (Bio-Rad Laboratories, Hercules, CA) or GenescreenPlus (DuPont Co., Wilmington, DE) membranes, which were neutralized in 2x SSC, UV-crosslinked for 1 min and prehybridized at 65°C overnight. Hybridization to probes labeled by the hexanucleotide technique (Prime-It II random primer kit, Stratagene, La Jolla, CA) with [α-^{32}P]dCTP (Amersham, Arlington Heights, IL) was performed at 65°C overnight. Membranes were washed in a solution containing 2x SSC, 0.1% SDS and 1 mM EDTA for 30 min at 65°C and then in a solution containing 0.5x SSC, 0.1% SDS and 1 mM EDTA for 5-15 min at 65°C. Blots were exposed to X-ray film (DuPont) in FisherBiotech (Pittsburgh, PA) autoradiography cassettes for 1-7 days and films were developed using a Konica QX60A processor.

Northern Blot Analysis

Isolation of poly(A)$^+$ RNA from fibroblast and Epstein-Barr virus-transformed B lymphoblastoid cell lines, formaldehyde agarose gel electrophoresis and northern blotting to ZetaProbe membranes were performed according to standard procedures (Sambrook *et*

al., 1989). Northern blots of 2 μg poly(A)$^+$ RNA from various mouse tissues, human cancer cell lines and human lymphoid tissues, as well as a dot-blot containing poly(A)$^+$ RNA from 50 human adult and fetal tissues were obtained from Clontech Laboratories. The quantity of poly(A)$^+$ RNA on each dot was normalized for 8 housekeeping genes, and varied from 100 to 500 ng. Membranes were hybridized with various probes labeled with [α-^{32}P]dCTP by the hexanucleotide technique (Prime-It II random primer kit). Hybridization conditions were the same as those employed with Southern blots. Autoradiography or phosphorimaging was performed using X-ray film or phosphorimager (Molecular Dynamics, Sunnyvale, CA), respectively.

Simple Sequence Length Polymorphism (SSLP) PCR Amplification

PCR reactions with Taq DNA polymerase (Boehringer Mannheim) were performed using 40 ng of mouse genomic DNA, 1 μmol/L of each primer (Research Genetics, Inc., Huntsville, AL) and 200 μmol/L of each dNTP in a 20 μL reaction. PCR conditions were: 95°C for 2 min, then 30-36 cycles of 94°C for 20 s, 55 °C for 30 s, 72 °C for 20 s, followed by 72 °C for 1 min. Amplification products were separated on 3% agarose gels, visualized by ethidium bromide staining and photographed.

Backcross Mouse Panel and Genetic Mapping

C57BL/6J-*bg*^{J} X (C57BL/6J-*bg*^{J} x CAST/EiJ)F_1 backcross mice were bred and maintained as described in Holcombe *et al.*, 1991. Genomic DNA was isolated from

mouse organs using standard techniques (Sambrook *et al.*, 1989) and Southern blots were prepared. The chromosomal assignment of each probe was determined by linkage analysis in 93 intersubspecific backcross mice. The 192 genetic markers (156 microsatellite markers, 14 genes, 20 expressed sequence tags (Tchernev *et al.*, 1997) and 2 coat color loci) that have been previously mapped in this panel served as anchor loci with average genetic distance between them ~ 10 cM. Gene order was determined by minimization of crossover events and elimination of double crossovers between linked genes. Intergene distances were expressed in centimorgans (cM) and standard errors were determined as previously described (Green, 1981; Bishop, 1985).

Cloning of *LYST* and *LYST2* Bait cDNA Fragments in Yeast Two-Hybrid Vectors

Fourteen overlapping cDNA fragments (with average size of ~850 bp each), covering the entire coding domain of *LYST*, were PCR amplified from *LYST* cDNAs cloned in the pCR 2.1 vector. *LYST2* cDNA fragments were amplified from EST clone #32273, obtained from Image Consortium. Primer design ensured that all cDNA fragments would be amplified in frame and that known domains with potential functional significance would be preserved. All primers contained specific restriction sites at their 5'-ends and all reverse primers contained stop codons near their 5'-ends. The sequence of all forward and reverse primers with their restriction sites, and the length and the position (according to GenBank accession number U67615 for *LYST*) of the amplified cDNA fragments are shown in Table 2 (for *LYST*) and in Table 3 (for *LYST2*).

Table 2. PCR primers, with restriction sites on their 5'-ends, used to amplify bait cDNA fragments from the *LYST* gene

GENE, FRAGMENT #	PRIMER SEQUENCE, 5'→ 3' (forward – upper sequence; reverse – lower sequence)	RESTR. ENZYME	FRAGMENT POSITION,bp	FRAGM SIZE,bp
LYST 1	GCGGTGGATCCCATGAGCACCGACAGTAACTCAC GCGGAATTCTCATAGTGTGGGCACTACACTGG	*Bam*HI *Eco*RI	190-1056	866
LYST 2	GCGGTGGATCCCACACTAACTGAGTTCCTAGCAGGCTTTGGGGACTGC $(GAGAGA)_3$CTCGAGTCAATCCATACAACAGCATATTCCAATG	*Bam*HI *Xho*I	1051-1950	899
LYST 3	GCGGAACTAGTGGATCCCAAATCTGTAATCATTCC GGAATTCTCACTGAACAACTATATTGCCTTTCTG	*Spe*I *Eco*RI	1948-2370	422
LYST 4	GCGAGTGTGGATCCCCAGAAAGGCAATATAGTTGTTCAG GCGGAATTCTCACTCCTCTTTGTGACTTCTGAAC	*Bam*HI *Eco*RI	2347-3213	866
LYST 5	GCGGTGGATCCCCTGTTCAGAAGTCACAAAGAGG GGAATTCTCAAGCAGAAAGCAAATTTAATTCCAG	*Bam*HI *Eco*RI	3190-4032	842
LYST 6	GCGGTGGATCCCCTGGAATTAAATTTGCTTTCTGC GCGGAATTCTCAGGAGCCCAGTGAAATTATATG	*Bam*HI *Eco*RI	4009-4821	812
LYST 7	GCGGTGGATCCCTCCAAAGCGTTGATGATCCAAG GCGGAATTCTCAAACTCGCAGTGCTAATGCTTG	*Bam*HI *Eco*RI	4819-5700	881
LYST 8	GCGGTGGATCCCACTCAAGCATTAGCACTGCGAG $(GAGAGA)_2$CTCGAGTCAGGCAACATAAGTATCTGCAATATTTTG	*Bam*HI *Xho*I	5677-6612	935
LYST 9	GCGGTGGATCCCCAAAATATTGCAGATACTTATGTTGCC GCGGAATTCTCATCCCATGTTTCTCACATCTTCCAG	*Bam*HI *Eco*RI	6586-7449	863
LYST 10	GCGGTGGATCCCCTGGAAGATGTGAGAAACATGGG GCGGAATTCTCAGGTCTGGAAAACTGAGGTCTTG	*Bam*HI *Eco*RI	7426-8238	812
LYST 11	GCGGTGGATCCCAAGACCTCAGTTTTCCAGACCG GCGGAATTCTCAATCCAACTGCCATGAGGTTGG	*Bam*HI *Eco*RI	8218-9039	821
LYST 12	GCGGTGGATCCCGATCCAACAGAAGGGCCAAATC GCGGAATTCTCATGTCAGAGCGGTGATGTTACC	*Bam*HI *Eco*RI	9037-9585	548
LYST 13	GCGGTGGATCCCAAGGTTCGTGATGATGTATACCAC GCGGAATTCTCACAAGCCTTTTATCCATGACAAAGG	*Bam*HI *Eco*RI	9502-10590	1088
LYST 14	GCGGTGGATCCCTGGATAAAAGGCTTGAAATGGGG GCGGAATTCTCATGAAGTTCATTCGCATTCACCC	*Bam*HI *Eco*RI	10576-11611	1035

Table 3. PCR primers, with restriction sites on their 5'-ends, used to amplify bait cDNA fragments from the *LYST2* gene

GENE, FRAGMENT #	PRIMER SEQUENCE, 5'→ 3' (forward – upper sequence; reverse – lower sequence)	RESTR. ENZYME	FRAGMENT POSITION, bp*	FRAGM. SIZE, bp
LYST2 *1*	GCGGAACTAGTGACTTCTGATGTAAAGGAAC GCGGAATTCTCAATAGCGATTATCTGCGTGTAC	*Spe*I *Eco*RI	(3-794)	791
LYST2 *2*	GCGGAACTAGTGACTTCTGATGTAAAGGAAC GCGGAATTCTCATGGATCCATTTCAATGGG	*Spe*I *Eco*RI	(3-686)	683
LYST2 *3*	GCGGAACTAGTGGTAACAGCAGATAATCGCTAT GCGGAATTCTCACCAGGGTTAAATGTACAGTTG	*Spe*I *Eco*RI	(774-1424)	650
LYST2 *4*	GCGGAACTAGTGACTTCTGATGTAAAGGAAC GCGGAATTCTCACCAGGGTTAAATGTACAGTTG	*Spe*I *Eco*RI	(3-1424)	1421

* Since the entire *LYST2* cDNA sequence is not known, fragment positions are relative to the beginning of the known *LYST2* sequence.

KlenTaq DNA polymerase and 12.5 pmol of each primer were used in each 50 μL amplification reaction. PCR conditions were: 94°C for 5 min, then 30 cycles of 94 °C for 40 s, 50 °C for 40 s, 72 °C for 2 min, followed by 72 °C for 5 min. PCR products were visualized on ethidium bromide-stained agarose gels, purified using QIAquick PCR purification kit (QIAGEN Inc., Valencia, CA) and digested with the corresponding restriction enzymes (Tables 2 and 3) in 20 μL-reactions overnight at 37 °C. Digested PCR products were excised from 1% low melting point agarose gels, purified using QIAquick gel extraction kit (QIAGEN Inc.) and ligated in the yeast Gal4 activation domain cloning vector pAD-GAL4 (modified pGAD GH, Clontech Laboratories) and in the yeast Gal4 DNA-binding domain vector pGB-GAL4 (modified pGBT9, Clontech Laboratories). DH10B competent cells were transformed with the purified constructs by electroporation, incubated for 1 h at 37 °C with shaking (RPM=150) and plated on LB/ampicillin (50 μg/mL) plates containing X-Gal and IPTG. The identity of cloned fragments was confirmed by PCR of bacterial colonies with one vector primer and one fragment-specific primer under the same amplification conditions as above. LB/ampicillin liquid cultures were inoculated with positive colonies and incubated overnight at 37 °C with shaking. Purified plasmids were isolated using QIAprep Spin Plasmid kit (QIAGEN Inc.) and visualized on ethidium bromide-stained agarose gels. All inserts were sequenced to ensure that PCR amplification reproduced an accurate copy of the *LYST* and *LYST2* sequence and that the cDNA fragments were cloned in frame.

Yeast Two-Hybrid Screens

Two types of screening for protein-protein interactions were performed by the yeast two-hybrid method. In a "forward" screen, one hybrid consisted of the DNA binding domain of the yeast transcriptional activator Gal4 fused to a "bait" portion of LYST or LYST2. The other hybrid consisted of the Gal4 activation domain fused to "prey" sequences encoded by a mammalian cDNA library. In a "reverse" screen, the "bait" part of LYST or LYST2 was fused to the Gal4 activation domain, and the "prey" sequences of the mammalian cDNA library were fused to the Gal4 DNA binding domain. The prey cDNAs were obtained from a human fetal brain cDNA library of 1×10^7 independent isolates (Clontech Laboratories). The library was synthesized from *Xho*I-$(dT)_{15}$-primed fetal brain mRNA (pooled from five male/female 19-22 week fetuses) that was directionally cloned into either pAD-GAL4, an yeast Gal4 activation domain cloning vector including the *LEU2* gene for selection in yeast deficient in leucine biosynthesis, or pBD-GAL4, an yeast Gal4 DNA-binding domain cloning vector including the *TRP1* gene for selection in yeast deficient in tryptophane biosynthesis.

To ensure that the bait DNA-binding domain fusions do not possess intrinsic transcriptional activity, a test for "bait self-activation" was performed. YULH yeast cells, transformed by the lithium acetate/polyethylene glycol method (Ito *et al.*, 1983) with the pBD-GAL4 bait constructs, were grown at 30°C on SC-Trp medium to select for the presence of the DNA-binding domain plasmid. The selected yeast cells were then plated on SC-Ura medium, grown overnight at 30°C, and examined for growth. Cells that did not grow contained baits which were not "self-activating" proteins, that is, these proteins

required interaction with a second protein domain in order to form a functional complex that can activate the transcription of reporter genes. Baits from non-growing cells were used for further screening.

In the forward screens, the constructs encoding the baits were introduced by lithium acetate/polyethylene glycol transformation into the yeast strain YULH (mating type a, *ura3*, *his3*, *lys2*, *Ade2*, *trp1*, *leu2*, *gal4*, *gal80*, *GAL1-URA3*, *GAL1-lacZ*), while the prey sequences were transformed into the yeast strain N106r (mating type α, *ura3*, *his3*, *ade2*, *trp1*, *leu2*, *gal4*, *gal80*, *cyh*^{r}, *Lys2::GAL1*$_{UAS}$*-HIS3*$_{TATA}$*-HIS3*, *ura3::GAL1*$_{UAS}$*-GAL*$_{TATA}$*-lacZ*). For the reverse screens, baits were transformed into N106r and preys into YULH. Initially, a limited screening using approximately 50,000 prey library members was performed to ensure that bait fusions do not interact promiscuously with many proteins (test for non-specific binding). Only baits that did not possess intrinsic transcriptional activity (*i.e.*, passed the test for "bait self-activation") and that did not interact indiscriminately with numerous proteins (*i.e.*, passed the test for non-specific binding), were used for screening the entire prey library. In these full-scale screens, the two transformed populations were mated using standard methods (Sherman *et al.*, 1991). Briefly, cells were grown at 30°C until mid-to-late log phase on media that selected for the presence of the appropriate plasmids, *i.e.* yeast containing activation domain constructs were grown on SC-Leu medium, while cells containing DNA binding domain plasmids were grown on SC-Trp medium. The two mating strains, α and a, were then diluted in YPAD media (Sherman *et al.*, 1991), filtered onto nitrocellulose membranes, and incubated at 30°C for 6-8 hours. The cells were then transferred to media selective for

the desired diploids, *i.e.*, for yeast harboring reporter genes for beta-galactosidase, uracil auxotrophy, and histidine auxotrophy, and for expression of the vectors encoding the bait and prey. The mating products were plated on SC (synthetic complete) media (Kaiser *et al.*, 1994) lacking adenine and lysine (to select for successful mating), leucine and tryptophane (to select for expression of genes encoded by the bait and prey plasmids), and uracil and histidine (to select for protein interactions). This medium is referred to as SCS medium, for SC Selective medium.

Selected clones were tested for expression of β-galactosidase to confirm the formation of an LYST:LYST-interacting protein (IP), resp. LYST2:LYST2-IP, interactions. Filter-lift β-galactosidase assays were performed as modified from the protocol of Breeden and Nasmyth, 1985. Colonies were patched onto SCS plates, grown overnight, and replicated onto Whatman No.1 filters. The filters were then assayed for β-galactosidase activity. Colonies that were positive turned a visible blue, indicating that a protein-protein interaction has occurred, enabling the transcription of the reporter gene *LacZ*.

Cells in colonies positive for protein interaction contained both DNA-binding and activation-domain plasmids. These cells were regrown as single isolates in individual wells of 96-well plates. Ten microliters of each isolate were lysed, the inserts within the pAD-GAL4 and pBD-GAL4 plasmids were amplified by PCR using primers specific for the insert-flanking sequences of the vectors, and approximately 300 amino-terminal bases of each insert were determined using an ABI 377 sequencer. Comparison to known sequences was made using the BLAST program publicly available through the National

Center for Biotechnology Information (N.C.B.I.). The assembly and identity searches of the sequences encoding human EST were performed by using publicly available EST assembly databases such as the N.C.B.I. BlastN 2.0 program (Altschul *et al.*, 1990). Sequences that aligned with 95% or greater identity at the nucleic acid level over their termini of at least 30 nucleotides were utilized if the alignment resulted in 5' extension or 3' extension of the EST sequence. Once this first assembly was complete, the extended sequence was again subject to the BlastN comparison to detect new homologies to the added extensions. The sequence was extended in both directions until new related sequences that allowed extension of the assembled sequence were no longer detected. The assembled EST sequence was subjected to further searches using the BlastX 2.0 program for the identification of protein coding regions by database similarity search (Gish and States, 1993). The BlastX software translates the DNA sequence in all six reading frames and compares the translated protein sequence with those in protein databases. The statistical significance is estimated under the assumption that the equivalent of one entire reading frame in the query sequence codes for protein and that significant alignments will involve only coding reading frames. Furthermore, the sequences were analyzed for open reading frames using software that translates the assembled DNA sequence in all six reading frames using the standard genetic code. The interacting ESTs were obtained from directionally-cloned libraries, and thus the direction of translation of the assembled EST is known as 5' to 3' and the open reading frame is also known. ORFs longer than 50 amino acids following an initiator codon or an ORF with no initiator methionine encoded at the 5'-end were determined to be possible protein products, and were compared to

sequences in protein data bases using the BLASTP 2.0 program (Altschul *et al.*, 1990). Further protein sequence analysis was performed after selecting a suitable open reading frame. The protein sequence was compared to previously characterized protein domains present in the BLOCKS and PRODOM motif databases (Bairoch, 1992; Henikoff and Henikoff, 1991; Nakai and Kanehisa, 1992; Wallace and Henikoff, 1992).

Confirmation of the Specificity of Protein Interactions by the Yeast Two-Hybrid System

In addition to sequencing all baits, and the tests for "bait self-activation" and for non-specific binding, described in the previous section, several other tests were performed to determine the specificity of the detected bait:prey interactions. In the test for "prey self-activation", YULH yeast cells were transformed with individual plasmids encoding the DNA-binding domain fusions of all preys that were detected to interact with LYST or LYST2. The transformed haploid yeast cells were grown at 30°C on SC-Trp medium to select for the presence of the DNA-binding domain plasmid. The selected yeast cells were plated on SC-Ura medium, grown overnight at 30°C, and examined for growth. Cells that did not grow contained preys which were not "self-activating" proteins, that is, these proteins required interaction with a second protein domain in order to form a functional complex that can activate the transcription of reporter genes. Preys from non-growing cells were used in the other confirmation tests.

YULH and N106r haploid yeast cells, transformed with DNA-binding domain, resp. activation domain fusions of preys that were detected to interact with LYST or LYST2, were grown on SC-Trp or SC-Leu media to select for the corresponding

plasmids. Filter–lift β-galactosidase assay was performed on the selected haploid cells. Colonies that did not turn blue contained preys that needed to interact with another protein in order to activate the transcription of the *LacZ* reporter gene. Only cells that did not grow on SC-Ura medium (*i.e.*, passed the test for "prey self-activation") and that did not turn blue (*i.e.*, passed the β-galactosidase test for haploids) were used in the next confirmation tests.

The following "matrix mating" test was performed to duplicate the initial forward and reverse screens and to determine whether the same protein-protein interactions would be detected. YULH yeast cells, transformed with DNA-binding domain fusions of baits (forward screen) or preys (reverse screen), and N106r cells, transformed with activation domain fusions of baits (reverse screen) or preys (forward screen), were grown in YPAD medium at 30°C. The "matrix mating" was performed by spreading yeast containing individual baits on YPAD plates, transferring (in 96-well format) prey-containing cells onto the spread baits, and incubating the plates overnight at 30°C. Each plate contained also positive (proteins that are known to interact) and negative (proteins that do not interact with a particular bait) controls. Resulting diploid colonies were replica-plated on 2 types of media: on SCS, which selected for successful mating, for protein-protein interactions and for presence of plasmids, and on SC-Leu-Trp plates, which were used for a filter-lift β-galactosidase assay. Only baits and preys from diploid yeast cells that were able to grow on SCS medium (*i.e.*, passed the "matrix mating" test) and that turned blue (*i.e.*, passed the β-galactosidase test for diploids), were PCR amplified with vector primers and the resulting PCR products were sequenced. If sequencing determined that

the identity of a particular bait:prey pair matched that obtained after the initial screens, the interaction was considered confirmed by the yeast two-hybrid method. An overview of the described yeast two-hybrid screening and confirmation procedures is shown in Figure 2.

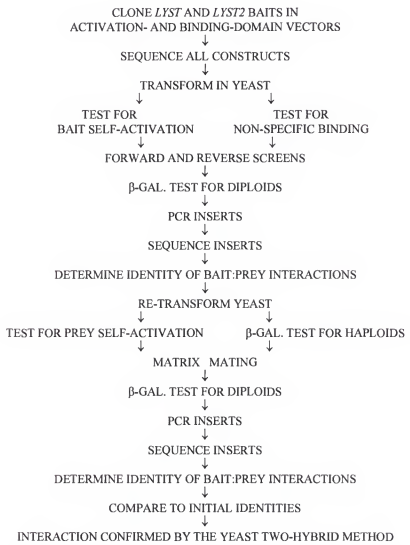

Figure 2. Overview of the screening and confirmation procedures by the yeast two-hybrid method.

RESULTS

Identification of Mutations in Patients with Chediak-Higashi Syndrome and in Beige Mice

Detection of mutations in the *CHS* gene of human patients and beige mice was undertaken for three main reasons:

- To provide additional evidence that *LYST*, resp. *Lyst* is indeed the gene responsible for the Chediak-Higashi syndrome in human and for the beige mutation in mouse
- To identify critical amino acids and domains that are essential to the function of the CHS gene
- To determine whether a genotype-phenotype correlation could be established between different genetic defects and specific disease manifestations.

A summary of mutations detected in patients with Chediak-Higashi syndrome is shown in Table 4. Several different approaches were used for mutation detection. As an initial screen for mutations, we analyzed northern blots of poly(A)$^+$ RNA from CHS patients. The largest mRNA species (~13 kb) was greatly reduced in abundance or absent in lymphoblastoid mRNA of patients P1 and P3, respectively (Fig. 3A), while the smaller transcript (~4.4 kb) was present and undiminished in abundance in all three patients.

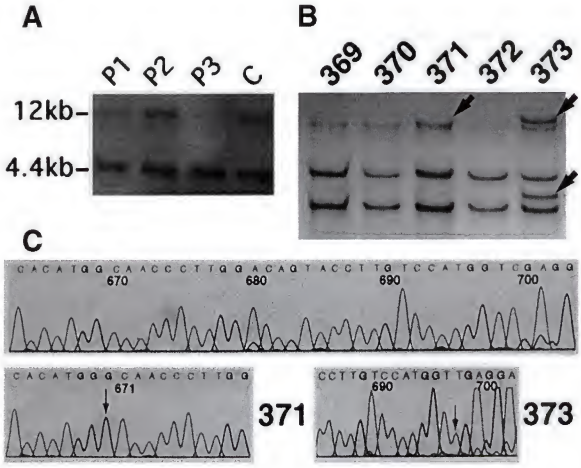

Figure 3. Mutation analysis in CHS patients. (**A**) Northern blot of 2 μg aliquots of lymphoblastoid poly(A)$^+$ RNA from CHS patients and a control. The probe used for hybridization corresponds to nucleotides 490-817 of the *CHS* cDNA (GenBank accession No. U70064). Exposure time for autoradiography was 48 h. (**B**) SSCP analysis of cDNA corresponding to nucleotides 439-806 of the *CHS* cDNA. Each lane contains samples from individual patients as indicated. Extra bands in lanes corresponding to patients 371 and 373 are denoted with an arrow. (**C**) Sequence chromatograms showing mutations in cDNA clones from patients 371 and 373. The upper part is normal human *CHS* cDNA sequence. The arrows indicate the positions of a G insertion (coding domain nucleotide 118, patient 371), and C→T substitution (coding domain nucleotide 148, patient 373). The sense strand is shown.

While the selective loss of the larger transcript in these two patients cannot be excluded definitively, rehybridization of this blot with an actin probe suggested that the absence of the larger transcript was not due to uneven gel loading or RNA degradation. Fibroblast poly(A)$^+$ RNA from three other CHS patients (369, 370 and 373) showed a moderate reduction in the ~13 kb mRNA (51-60% of control by densitometry), while the ~4.4 kb mRNA was essentially unaltered in abundance (103-147% of control).

Single-strand conformation polymorphism (SSCP) analysis was undertaken using cDNA samples derived from lymphoblastoid or fibroblast cell lines from CHS patients, in collaboration with Drs. Steve Colman and Margaret Wallace. Anomalous bands were detected in PCR products from the 5'-end of the open reading frame in two unrelated CHS patients (371 and 373) that were different from those with probable selective loss of the larger transcript on northern blots (Fig. 3B). Sequence analysis identified a C→T substitution at nucleotide 148 of the coding domain in patient 373 (Fig.3C). Four of the nine cDNA clones derived from patient 373 contained this mutation. Restriction enzyme digestion confirmed the mutation: *Taq*I digestion of cDNA (nucleotides 520-808) showed loss of this restriction site in patient 373 to be heterozygous. The C→T substitution creates a stop codon at amino acid 50 (R50X). The mutation in patient 373 occurs within a CpG dinucleotide, which, when the cytosine is methylated, represents a common hotspot for mammalian mutations due to deamination.

In patient 371, a G insertion was found at nucleotide 118 of the coding domain, resulting in a frame shift at codon 40 and termination after amino acid 62 (Fig. 3C). While analysis of genomic DNA had shown patient 371 to be heterozygous for this

mutation (Nagle *et al.*, 1996), each of five cDNA clones isolated from lymphoblasts of this patient contained the G insertion. Allele-specific oligonucleotide hybridization of cDNA from this patient failed to detect a signal with an oligonucleotide corresponding to the normal allele, suggesting that the patient is transcriptionally hemizygous for the G insertion. Alternatively, the lack of allele-specific oligonucleotide hybridization may reflect a base mismatch (such as a polymorphism) within the patient's cDNA.

Mutations were identified in three other CHS patients by cloning and sequencing fragments of their *LYST* coding domains. A homozygous C→T substitution at nucleotide 3085 of the coding domain creating a stop codon at amino acid 1029 (Q1029X) was found in patient 370. In patient 369, a heterozygous frameshift mutation was detected. Nucleotides 3073 and 3074 of the coding domain were deleted in two of five cDNA clones, resulting in a frameshift at codon 1026 and termination at codon 1030. A homozygous C→T substitution at nucleotide 3310 of the coding domain was identified in patient 372. This mutation created a stop codon at amino acid 1103, resulting in premature termination. Despite mutation analysis of >6 kb of open reading frame, the mutations in CHS patients P1 and P3, that may have a selective loss of the larger transcript on northern blots, have not been identified.

Lymphoblasts from all of these patients (369, 370, 371, 373, P1 and P3) contain the giant perinuclear lysosomal vesicles that are the hallmark of the Chediak-Higashi syndrome. The parents of patients 369 and 370 are known not to have been consanguineous. Patients 369, 370 and 371 had typical clinical presentations of CHS, with recurrent childhood infections and oculocutaneous albinism. In contrast, the clinical

course of patient 373 was milder – this patient has not had systemic infections and remains alive at age 37. Patient 373 does, however, have hypopigmented hair and irides, as well as peripheral neuropathy. A summary of the described mutations, detected in patients with Chediak-Higashi syndrome, is shown in Table 4.

Table 4. Mutations identified in patients with Chediak-Higashi syndrome.

PATIENT	ZYGOSITY	MUTATION	CONSEQUENCE
369	Heterozygous	bp 3073 & 3074 deleted	Codon 1026 frameshift, stop at codon 1030
370	Homozygous	C3085T	Q1029X
371	Heterozygous	G insertion at bp 118	Codon 40 frameshift, stop at codon 62
372	Homozygous	C3310T	stop at codon 1103
373	Heterozygous	C148T	R50X
X Nagle *et al.*, 1996	Homozygous	bp 1467 deleted	Codon 489 frameshift, stop at codon 566
P1	?	↓↓ 13.5 kb mRNA	lack of protein
P3	?	↓↓ 13.5 kb mRNA	lack of protein

Mutations were also identified in several beige alleles. A 5-kb genomic deletion that contained the 3'-end of *Lyst* exon β, and exons γ and δ, was found in bg^{1J} DNA. This mutation was detected by hybridization of *Lyst*-derived probes to Southern blots of genomic DNA from all potential progenitor mouse strains (Fig. 4), and by PCR amplification of several *Lyst* fragments (Fig. 5). This deletion corresponds to a loss of approximately 400 internal amino acids of the predicted Lyst protein. Furthermore, whereas the 5'-end of the bg^{1J} deletion occurs within *Lyst* exon β, the 3'-end is intronic. Therefore the truncated *Lyst* mRNA in bg^{1J} mice is also anticipated to splice incorrectly, terminate prematurely, and lack polyadenylation. Unfortunately, it was not possible to confirm this prediction since bg^{1J} mice are extinct and only archived genomic DNA was available.

Quantitative reverse transcription (RT)-PCR demonstrated a moderate reduction in *Lyst* mRNA in *bg* and bg^{J} liver, and a gross reduction in bg^{2J} (*Lyst* ΔOD after normalization for β-actin mRNA: +/+, 1.00; bg^{2J}/bg^{2J}, 0.19; *bg/bg*, 0.28; bg^{J}/bg^{J}, 0.40) (Fig. 6). A commensurate reduction in bg^{2J} transcript abundance was noted by using several primer pairs derived from different regions of the *Lyst* cDNA. Aberrant *Lyst* RT-PCR products were not observed. The particularly striking (more than fivefold) reduction in *Lyst* expression evident in bg^{2J} homozygotes suggested the existence of a mutation in bg^{2J} that results in decreased transcription or mRNA instability. The molecular basis of the reduction in *Lyst* mRNA in bg^{2J} is not yet known.

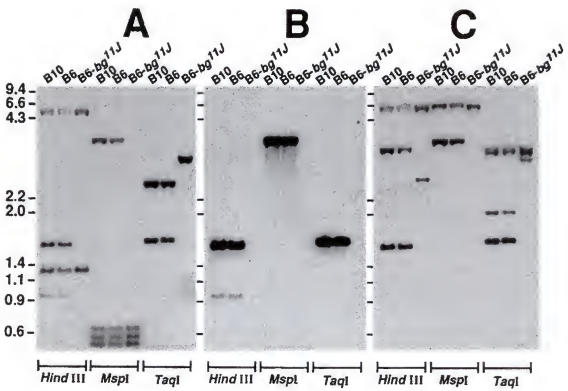

Figure 4. Southern blot identification of an intragenic *Lyst* deletion in bg^{11J}. A Southern blot was sequentially hybridized with 3 *Lyst* probes: (**A**) the probe (nucleotides 1262-3433 of *Lyst* cDNA) extends upstream of the bg^{11J} deletion. (**B**) the probe (nucleotides 2835-3433 of *Lyst* cDNA) is completely deleted; (**C**) the probe (nucleotides 3594-4237 of *Lyst* cDNA) extends downstream of the bg^{11J} deletion. Restriction endonucleases are indicated at the bottom of each panel, and molecular size standards (in kb) are shown to the left. Southern blots were prepared from genomic DNA of all potential progenitor mouse strains, but only C57BL/10J, C57BL/6J and C57BL/6J- bg^{11J} are shown.

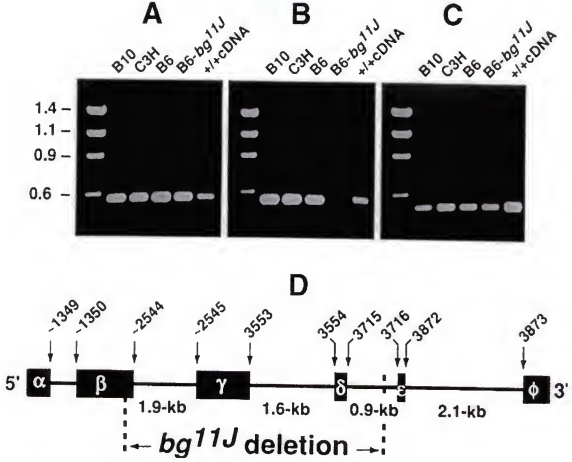

Figure 5. PCR analysis of the bg^{11J} deletion and genomic structure of *Lyst* in the vicinity of the deletion. C57BL/10J, C3HeB/FeJ, C57BL/6J and C57BL/6J- bg^{11J} genomic DNA and *Lyst* cDNA were used templates in the PCR reactions. Amplicons illustrated correspond to: (**A**) *Lyst* cDNA nucleotides 1337-1837, which represent exon β and are upstream of the deletion; (**B**) nucleotides 2670-3210, which represent exon γ, deleted in bg^{11J} DNA; (**C**) nucleotides 4913-5433, which represent an exon downstream from the deletion. No amplicon was observed in control PCR reactions performed without template. More than 30 other STSs that had been localized within the *bg* non-recombinant interval amplified normally from bg^{11J} DNA. (**D**) Genomic structure of *Lyst* in the vicinity of the bg^{11J} deletion. *Lyst* exons (α, β, γ, δ, ε and ϕ) are depicted by black boxes, and intervening introns by a solid line. Nucleotides of the mouse *Lyst* cDNA that correspond to exonic boundaries are indicated above the boxes. The 3' end of exon β, and all of exons γ and δ, are deleted in bg^{11J} DNA. Genomic structure and intronic sequences were ascertained by sequence analysis of nested PCR products, performed with exonic primers and P1 genomic clone as template. Boundaries of the bg^{11J} deletion were determined by PCR of genomic DNA.

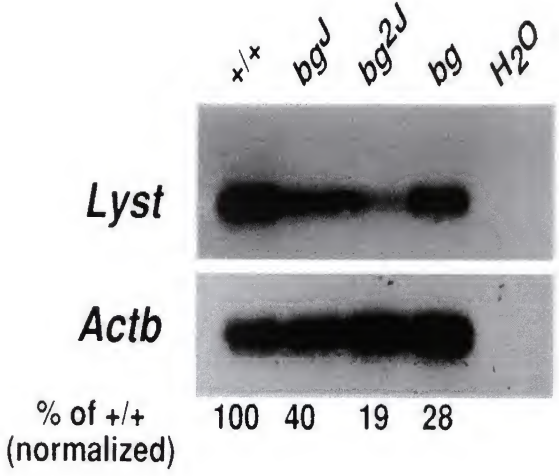

Figure 6. Quantitative reverse transcription (RT)-PCR analysis of *Lyst* in several *bg* alleles. *Lyst* mRNA from bg^{J}, bg^{2J}, *bg* and control livers was reverse transcribed, PCR amplified and the amount of the resulting PCR products was quantitated by measuring ΔOD after normalization for β-actin mRNA (% ΔOD for *Lyst* are shown below the gel image).

Identification and Characterization of Lyst mRNA Isoforms

The investigation of mRNA isoforms of the *CHS* gene was undertaken for the following reasons:

- To resolve the discrepancy between the initial reports about the identification of the beige gene. The partial cDNA sequence, published by Perou *et al.*, 1996b, did not match the complete open reading frame, reported by Barbosa *et al.*, 1996.
- To identify the isoform of primary functional significance, as well as other alternatively spliced variants that may have some importance.
- To determine which of the predicted domains are crucial for the function of the CHS protein.

The existence of more than one mRNA isoform of the mouse beige gene (designated *Lyst*) had been suggested by northern analysis (Fig. 7) and by the two different *Lyst* cDNA sequences that were reported. Northern blots indicated that *Lyst* is ubiquitously transcribed, both temporally and spatially, in mouse and human tissues. Complex alternative splicing, with both constitutive and anatomically restricted *Lyst* mRNA isoforms, was also revealed. The largest *Lyst* transcript in human and mouse was 12-13 kb and additional smaller alternatively spliced variants were also present with varying abundance in different tissues (Fig. 7).

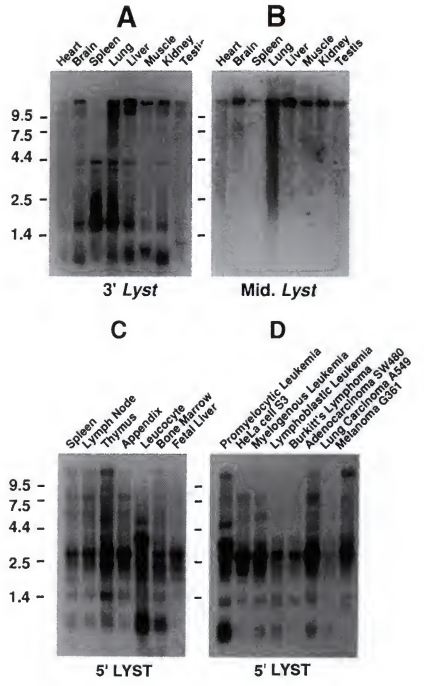

Figure 7. Northern blot analysis of mouse and human *LYST*. (**A**), (**B**): northern blots of 2 μg poly(A)$^+$ mRNA from various mouse tissues hybridized with probes that correspond to (A), nucleotides 4423-4631, and (B), nucleotides 1430-2457 (exon β) of mouse *Lyst* cDNA. (**C**), (**D**): northern blots of 2 μg poly(A)$^+$ mRNA from (C), human lymphoid tissues, and (D), human cancer cell lines, hybridized with a probe that corresponds to nucleotides 357-800 of human *LYST* cDNA. Molecular size standards (in kb) are shown to the left.

Mouse *Lyst* cDNA isoforms were identified by anchored nested PCR (3'-RACE-PCR). Two fragments (1.25 kb and 2 kb) were amplified from mouse spleen cDNA using this technique. The 1.25 kb clone contained the 3'-end of the previously described 5893 bp *Lyst* cDNA that corresponds to a small mRNA isoform (*Lyst*-II). The 5'-end of the 2 kb clone, however, contained sequence derived from the *Lyst*-II cDNA, while the 3'-end sequence was from the largest isoform (*Lyst*-I, previously called BG) of the beige gene. Reverse transcription (RT)-PCR confirmed that nucleotides 1-4706 of *Lyst* are common to both mRNA isoforms (Fig. 8A). The large isoform cDNA that contains the entire coding domain was assembled from nucleotides 1-4706 of *Lyst*, the 2 kb 3'RACE-PCR clone, and 6824 nucleotides of BG cDNA. This 11 817 bp cDNA sequence (*Lyst*-I, GenBank accession number U70015) corresponds to the largest mRNA observed on northern blots (Fig. 7). This ~11.8 kb cDNA , however, is truncated at both the 5'-end of the 5'-untranslated region (UTR) and at the 3'-end of the 3'-UTR, and thus is smaller than the largest mRNA (~12-13 kb), observed on mouse northern blots.

Analysis of a P1 genomic clone (number 8592), containing the entire beige gene, revealed that the 11,817 bp *Lyst*-I cDNA results when splicing occurs from *Lyst* exon σ (containing nucleotide 4706) to downstream exon τ (Figs. 8 and 10). In contrast, incomplete splicing and reading through intron σ' (interposed between exons σ and τ) yields the truncated *Lyst*-II isoform, of length 5893 bp cDNA (*Lyst*-II, Fig. 8, GenBank accession number L77884). *Lyst* intron σ' encodes 37 in-frame amino acids followed by a stop codon and a polyadenylation signal. *Lyst*-II corresponds to one of several smaller mRNAs observed on northern blots (Fig. 7). Full-length cDNAs corresponding to the

Lyst-I and *Lyst*-II isoforms were both amplified from mouse bone marrow RNA by RT-PCR. The putative Lyst-I and Lyst-II proteins are of relative molecular mass 425 287 (M_r 425 kDa) and M_r 172.5 kDa, respectively.

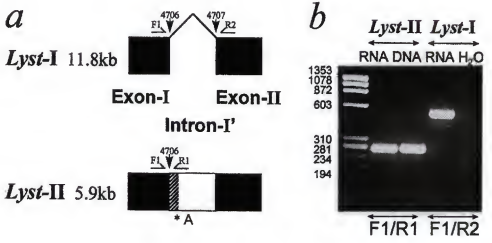

Figure 8. Alternative splicing of the CHS gene. (**A**) Alternative splicing of mouse *Lyst*. Solid lines represent *Lyst* exons σ and τ (not drawn to scale). Splicing of exon σ to exon τ occurs in the *Lyst*-I mRNA (~12 kb). The hatched box represents the intronic region that forms the 3' end of the *Lyst*-II ORF (5.9 kb). The intron contains a stop codon (*) and a polyadenylation signal (A). Nucleotide positions indicated are from GenBank accession No. L77884 (*Lyst*-II) and U70015 (*Lyst*-I). (**B**) Detection of mRNA isoforms of mouse *Lyst* by RT-PCR and genomic PCR. DNase-treated mouse melanocyte RNA was reverse transcribed and amplified with primers F1/R1 (expected amplicon size 273 bp) or F1/R2 (expected amplicon size 560 bp). RNase-treated C57BL/6J DNA was amplified with primers F1/R1.

cDNAs corresponding to the human homolog of the largest mRNA isoform of the beige gene were obtained by identification of human expressed sequence tags (ESTs) similar in sequence to mouse *Lyst*-I by database searches (GenBank accession numbers L77889, W26957 and H51623). Intervening cDNA sequences were isolated using RT-PCR, and a partial human cDNA sequence (GenBank accession number U70064; 7.1 kb) was assembled by alignment of these clones with the largest mouse *bg* gene cDNA. The predicted human and mouse peptides shared 82% identity over 1990 amino acids. The predicted human amino acid sequence contains a six amino acid insertion at residue 1039, relative to that of mouse. The complete sequence of the human CHS cDNA was reported by Nagle *et al.*, 1996. This 13.5 kb cDNA sequence corresponds to the largest mRNA observed on northern blots from human tissues (*CHS*-I, Fig. 7 and inset of Fig. 9). These northern blots also demonstrated the existence of a smaller human transcript (4.5 kb) that was similar in size to the small mouse *Lyst* mRNA, and that appeared to differ from the large isoform in distribution of expression in human tissues (Fig. 7). Assuming that the genomic derivation of the small human isoform might be the same as the small mouse isoform, the 3'-end of the small human isoform was identified by cloning human intron σ' using PCR of human genomic DNA with primers derived from exon σ and intron σ'. Similarly to intron σ' in the mouse, the sequence of the 5'-end of human intron σ' contained 17 codons in frame with exon σ, followed by a stop codon (*CHS*-II, inset of Fig. 9). cDNA corresponding to this short human isoform was amplified from human peripheral blood RNA by RT-PCR with primers from a 5' exon and from intron σ', indicating that this intron was indeed present in *CHS*-II mRNA.

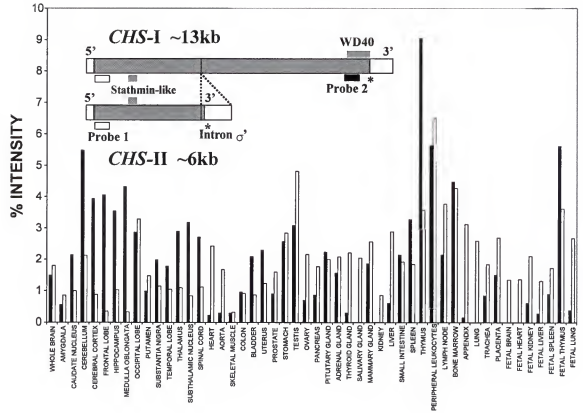

Figure 9. Comparison of the relative abundance of *CHS* (*LYST*) mRNA isoforms in 50 human tissues after normalization. Probe 2 (specific for the largest isoform *CHS*-I, or *LYST*-I) and the corresponding signal intensity levels are shown in black. Probe 1 (identifies all *CHS* transcripts) and the corresponding signal intensity levels are shown in white. Variation of the relative mRNA abundance among tissues was assessed by phosphorimaging, dot quantitation, subtraction of background and calculation of the intensity percentage for each tissue. The relative mRNA abundance between normalized tissues for each probe was expressed as the percentage of the total hybridization signal obtained for each tissue with that probe. The average percentage intensity for housekeeping genes in all tissues was ~2%. No hybridization was evident to control dots, containing yeast total RNA, yeast tRNA, *E. coli* rRNA, *E. coli* DNA, poly r(A), human $C_o t1$ DNA and human DNA. Exposure time for phosphorimaging was 20 h.

Nucleotides 1-5905 of human small and long isoform cDNAs are identical, and are followed either by intron σ' sequence in the short isoform (GenBank accession number U84744), or by exon τ and the rest of the exons, in the long isoform (inset of Fig. 9). The predicted intron-encoded amino termini of the short mouse and human isoform peptides share 65% identity.

Additional splicing complexity of smaller isoforms exists. The *Lyst*-III isoform lacks exons α and β, while in *Lyst*-IV, exons α, β and γ are absent. In both isoforms the alternative splicing occurs out of frame and results in termination shortly after the splice junction, suggesting that most likely these two isoforms do not have functional significance. These splice variants correspond to additional bands observed on northern blots (Fig. 7). A schematic representation of the four *Lyst* isoforms is shown in Fig. 10.

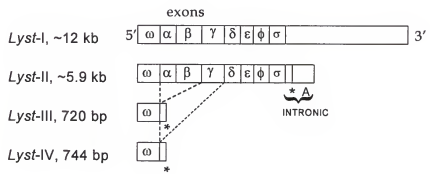

Figure 10. Schematic representation of the four mRNA isoforms of the Chediak-Higashi syndrome gene (*, stop codon; A, poly-A).

In an effort to evaluate the functional significance of the two largest *Lyst* isoforms, genetic complementation experiments were initiated in collaboration with Dr. Stephen Brandt, Vanderbilt University. Full-length *Lyst*-I and *Lyst*-II isoforms were amplified from mouse bone marrow cDNA by long-range PCR and cloned in expression vectors. The resulting constructs were expressed in cultured fibroblasts from beige mice, exhibiting the characteristic giant perinuclear granules in the cytoplasm. The effect was evaluated by observing the changes in the size and distribution of specifically stained lysosomes and endosomes. Some cells, expressing full-length *Lyst*-I, demonstrated more peripheral distribution of the stained granules in the cytoplasm and slight reduction of their size. Such changes were not visible in cells expressing the smaller *Lyst*-II isoform, supporting the hypothesis that the largest splice variant (*Lyst*-I) represents the mRNA isoform of primary functional significance. However, since the effects of the complementation were not very prominent and were not observed in all cells expressing *Lyst*-I, the significance of the described genetic complementation is questionable.

Analysis of northern blots of mouse mRNA had suggested that the relative abundance of the large and small transcripts differed from tissue to tissue (Fig. 7). The relative abundance of mRNA isoforms of the homologous gene in human tissues at different developmental stages was examined (Fig. 9) by sequential hybridization of a poly(A)$^+$ RNA dot-blot with a cDNA probe (inset of Fig. 9, Probe 2, shown in black, nucleotides 10941-11590 of human *CHS*, GenBank accession No U67615) specific for the largest isoform and with a cDNA probe (inset of Fig. 9, Probe 1, shown in white, nucleotides 190-445 of human *CHS*, GenBank accession No U70064 and U84744) that identifies all *CHS* transcripts (Fig. 9). The inset in Fig. 9 shows the human 5.8 kb cDNA

isoform (*CHS*-II, or *LYST*-II, which arises through incomplete splicing, with intron σ' supplying the 3'-end of the transcript), the largest isoform (*CHS*-I, or LYST-I, 13.5 kb, which results from removal of intron σ') and the location of the probes. The quantity of poly(A)$^+$ RNA on the blot was normalized to eight housekeeping genes (phospholipase, ribosomal protein S9, tubulin, highly basic 23 kDa protein, glyceraldehyde-3-phosphate dehydrogenase, hypoxantine guanine phosphoribosil transferase, β-actin and ubiquitin).

Using a probe that hybridized only to the largest mRNA isoform on northern blots, transcripts were most abundant in thymus (adult and fetal), peripheral blood leukocytes, bone marrow and several regions of the adult brain (Fig. 9). Interestingly, the largest mRNA isoform was not detected in fetal brain. There was also low relative expression of this isoform in heart, lung, kidney and liver in both adult and fetal tissues.

A somewhat different pattern of relative expression was evident upon rehybridization of the blot with a probe derived from the 5'-end of the coding domain of the CHS gene, a region that hybridized to all mRNA isoforms on northern blots. Consistent with the transcription pattern of the largest isoform, this probe detected abundant expression on peripheral blood leukocytes, thymus (adult and fetal) and bone marrow. However, several tissues with abundant large isoform transcripts exhibited considerably less relative expression with the 5'-probe, including most regions of the adult brain, fetal and adult thymus, and spleen. Furthermore, several tissues with negligible relative transcription of the large isoform exhibited increased relative expression with the 5'-probe, including adult and fetal heart, kidney, liver, and lung; adult aorta, thyroid gland, salivary gland and appendix; and fetal brain (Fig. 9).

Identification and Characterization of *Lyst2*, a Brain-Specific Member of the Chediak-Higashi Syndrome Gene Family

Identification and Sequence Analysis

In an effort to identify genes similar to the CHS gene, extensive database searches were performed. BLASTX comparison of mouse *Lyst* (U70015) to public EST databases revealed a significant, but non-identical, match with an uncharacterized human EST (R17955). The corresponding partial-length cDNA clone (#32273, length 1986 bp), derived from a human infant brain cDNA library, was sequenced. An additional 250 bp of cDNA sequence, located immediately 5' of clone 32273, was obtained by insert-vector PCR of a human fetal brain cDNA library. The resulting total sequence (2.2 kb, Fig. 11a) corresponded to the 3' end of the coding domain and the 3' UTR of a novel gene that was designated *LYST2* (lysosomal trafficking regulator 2, Fig.11b). Amino acid identity of *LYST2* with *LYST* varied from region to region: LYST2 amino acids 16-222 shared 50.2 % identity with LYST residues 3215-3425. This region corresponds to the BEACH domain (amino acids 3116-3461) of *LYST,* and exhibited sequence similarity to anonymous ORFs from *S. cerevisiae*, *C. elegans* and human CDC4L protein (Nagle *et al.*, 1996). In contrast, the WD repeat-containing carboxy-terminal domain of LYST, which is predicted to assume a six-bladed beta-propeller structure similar to the ß-subunit of heterotrimeric G proteins (Sondek et al., 1996), shared little amino acid similarity with LYST2. The latter region of LYST2 did, however, contain LYST-like WD repeats, and also exhibited significant sequence similarity to the ß-subunit of heterotrimeric G proteins

(30.4 % identity LYST2 amino acids 368-501 to Gß {P49027}). The *LYST* 3' UTR demonstrated 42.1% identity to 584 bp of the 3' UTR of *LYST2*.

In an effort to identify the mouse homolog of human *LYST2*, a mouse embryo cDNA library was screened with a human *LYST2* probe, and two overlapping clones of total length of ~2.5 kb were identified (Fig. 11c). The corresponding partial mouse putative protein (Lyst2, Fig. 11d) exhibited 29.4% identity over 676 amino acids to mouse Lyst and 96.2 % identity in 557 amino acids to human LYST2.

Genetic Mapping

In order to confirm that *Lyst* and *Lyst2* are indeed two different genes, and to evaluate the potential candidacy of *Lyst2* for mouse coat color mutations, *Lyst2* was mapped by cross-hybridization in the mouse using unique CAST/EiJ restriction fragment length polymorphisms with *Msp*I (~4.1 kb) and *Eco*RI (~3.6 kb). Linkage analysis using DNA from 93 intersubspecific backcross [C57BL/6J-*bg*^{J} X (C57BL/6J-*bg*^{J} x CAST/EiJ) F_1] mice revealed that *Lyst2* maps on mouse Chr 3 (Fig. 12). The best gene order and recombination frequency (± standard error) were: centromere - *D3Mit21* - 6.45±2.5 cM - *Lyst2* - 5.38±2.3 cM - *D3Mit22*-telomere (GenBank accession number AF072372).

LYST2 was mapped on human Chr 13 by hybridization to Southern blots of human-rodent somatic cell hybrids, in collaboration with Dr. Margaret Wallace. No cross-hybridization to *LYST* was observed (Fig.13). While the CHS gene maps on human Chr 1 and on mouse chromosome 13, LYST2 was localized to human Chr 13 and to mouse chromosome 3, confirming that LYST and LYST2 are indeed two different genes.

a) Human LYST2 DNA:

```
   1 CCGAAGAGAG CTGTGTTTTA TGCAGAGCGT TATGAGACAT GGGAAGATGA
  51 TCAAAGCCCA CCCTACCATT ATAATACCCA TTATTCAACA GCAACATATA
 101 CTTTATCCTG GCTTGTTCGA ATTGAACCTT TCACAACCTT CTTCCTCAAT
 151 GCAAATGATG GAAAATTTGA TCATCCAGAT CGAACCTTCT CATCCGTTGC
 201 AAGGTCTTGG AGAACTAGTC AGAGGGATAC TTCTGATGTA AAGGAACTAA
 251 TTCCAGAGTT CTACTACCTA CCAGAGATGT TTGTCAACAG TAATGGATAT
 301 AATCTTGGAG TCAGAGAAGA TGAAGTAGTG GTAAATGATG TTGATCTTCC
 351 CCCTTGGGCA AAAAAACCTG AAGACTTTGT GCGGATCAAC AGGATGGCCC
 401 TAGAAAGTGA ATTTGTTTCT TGCCAACTTC ATCAGTGGAT CGACCTTATA
 451 TTTGGCTATA AGCAGCGAGG ACCAGAAGCA GTTCGTGCTC TGAATGTTTT
 501 TCACTACTTG ACTTATGAAG GCTCTGTGAA CCTGGATAGT ATCACTGATC
 551 CTGTGCTCAG GGAGGCCATG GAGGCACAGA TACAGAACTT TGGACAGACG
 601 CCATCTCAGT TGCTTATTGA GCCACATCCG CCTCGGAACT CTGCCATGCA
 651 CCTGTGTTTC CTTCCACAGA GTCCGCTCAT GTTTAAAGAT CAGATGCAAC
 701 AGGATGTGAT AATGGTGCTG AAGTTTCCTT CAAATTCTCC AGTAACCCAT
 751 GTGGCAGCCA ACACTCTGCC CCACTTGACC ATCCCCGCAG TGGTGACAGT
 801 GACTTGCAGC CGACTCTTTG CAGTGAATAG ATGGCACAAC ACAGTAGGCC
 851 TCAGAGGAGC TCCAGGATAC TCCTTGGATC AAGCCCACCA TCTTCCCATT
 901 GAAATGGATC CATTAATAGC CAATAATTCA GGTGTAAACA AACGGCAGAT
 951 CACAGACCTC GTTGACCAGA GTATACAAAT CAATGCACAT TGTTTTGTGG
1001 TAACAGCAGA TAATCGCTAT ATTCTTATCT GTGGATTCTG GGATAAGAGC
1051 TTCAGAGTTT ATACTACAGA AACAGGGAAA TTGACTCAGA TTGTATTTGG
1101 CCATTGGGAT GTGGTCACTT GCTTGGCCAG GTCCGAGTCA TACATTGGTG
1151 GGGACTGCTA CATCGTGTCC GGATCTCGAG ATGCCACCCT GCTGCTCTGG
1201 TACTGGAGTG GGCGGCACCA TATCATAGGA GACAACCCTA ACAGCAGTGA
1251 CTATCCGGCA CCAAGAGCCG TCCTCACAGG CCATGACCAT GAAGTTGTCT
1301 GTGTTTCTGT CTGTGCAGAA CTTGGGCTTG TTATCAGTGG TGCTAAAGAG
1351 GGCCCTTGCC TTGTCCACAC CATCACTGGA GATTTGCTGA GAGCCCTTGA
1401 AGGACCAGAA AACTGCTTAT TCCCACGCTT GATATCTGTC TCCAGCGAAG
1451 GCCACTGTAT CATATACTAT GAACGAGGGC GATTCAGTAA TTTCAGCATT
1501 AATGGGAAAC TTTTGGCTCA AATGGAGATC AATGATTCAA CACGGGCCAT
1551 TCTCCTGAGC AGTGACGGCC AGAACCTGGT CACCGGAGGG GACAATGGGG
1601 TAGTAGAGGT CTGGCAGGCC TGTGACTTCA AGCAACTGTA CATTTAACCC
1651 TGGATGTGAT GCTGGCATTA GAGCAATGGA CTTGTCCCAT GACCAGAGGA
1701 CTCTGATCAC TGGCATGGCT TCTGGTAGCA TTGTAGCTTT TAATATAGAT
1751 TTTAATCGGT GGCATTATGA GCATCAGAAC AGATACTGAA GATAAAGGAA
1801 GAACCAAAAG CCAAGTTAAA GCTGAGGGCA CAAGTGCTGCATGGAAAGGC
1851 AATATCTCTG GTGGAAAAAA TTCGTCTACA TCGACCTCCG TTTGTACATT
1901 CCATCACACC CAGCAATAGC TGTACATTGT AGTCAGCAAC CATTTTACTT
1951 TGTGTGTTTT TTCACGACTG AACACCAGCT GCTATCAAGC AAGCTTATAT
2001 CATGTAAATT ATATGAATTA GGAGATGTTT TGGTAATTAT TTCATATATT
2051 GTTGTTTTATT GAGAAAAGGT TGTAGGATGT GTCACAAGAG ACTTTTGACA
2101 ATTCTGAGGA ACCTTGTGTC CAGTTGTTAC AAAGTTTAAG CTTTGAACCT
2151 AACCTGCATC CCATTTCCAG CCTCTTTTCA AGCTGAGAAA AAAAAAAAAA
2201 AAAAA
```

b) Human LYST2 peptide:

```
   1 PKRAVFYAER YETWEDDQSP PYHYNTHYST ATYTLSWLVR IEPFTTFFLN
  51 ANDGKFDHPD RTFSSVARSW RTSQRDTSDV KELIPEFYYL PEMFVNSNGY
 101 NLGVREDEVV VNDVDLPPWA KKPEDFVRIN RMALESEFVS CQLHQWIDLI
 151 FGYKQRGPEA VRALNVFHYL TYEGSVNLDS ITDPVLREAM EAQIQNFGQT
 201 PSQLLIEPHP PRNSAMHLCF LPQSPLMFKD QMQQDVIMVL KFPSNSPVTH
 251 VAANTLPHLT IPAVVTVTCS RLFAVNRWHN TVGLRGAPGY SLDQAHHLPI
 301 EMDPLIANNS GVNKRQITDL VDQSIQINAH CFVVTADNRY ILICGFWDKS
 351 FRVYTTETGK LTQIVFGHWD VVTCLARSES YIGGDCYIVS GSRDATLLLW
 401 YWSGRHHIIG DNPNSSDYPA PRAVLTGHDH EVVCVSVCAE LGLVISGAKE
 451 GPCLVHTITG DLLRALEGPE NCLFPRLISV SSEGHCIIYY ERGRFSNFSI
 501 NGKLLAQMEI NDSTRAILLS SDGQNLVTGG DNGVVEVWQA CDFKQLYI*
```

c) Mouse Lyst2 DNA:

```
   1 GCAGCAGGGC GAACCGGACC TCTGTGATGT TTAATTTTCC TGACCAAGCA
  51 ACAGTTAAAA AAGTTGTCTA CAGCTTGCCT CGGGTTGGAG TGGGGACCAG
 101 CTATGGTTTG CCACAAGCCA GGAGGATATC ACTGGCCACT CCTCGACAGC
 151 TGTATAAGTC TTCCAATATG ACTCAGCGCT GGCAAAGAAG GGAAATCTCC
 201 AACTTTGAGT ATTTGATGTT TCTCAACACG ATAGCAGGTC GGACGTATAA
 251 TGATCTGAAC CAGTATCCTG TGTTTCCATG GGTGTTAACA AACTATGAAT
```

```
 301 CAGAGGAGTT GGACCTGACT CTCCCAGGAA ACTTCAGGCA TCTGTCAAAG
 351 CCAAAAGGTG CTTTGAACCC GAAGAGAGCA GTGTTTTACG CAGAGCGCTA
 401 TGAGACATGG GAGGAGGATC AAAGCCCACC CTTCCACTAC AACACACATT
 451 ACTCAACGGC GACTTCCCCC CTTTCATGGC TTGTTCGGAT TGAGCCATTC
 501 ACAACCTTCT TCCTCAATGC AAATGATGGG AAATTTGACC ATCCAGACCG
 551 AACCTTCTCA TCCATTGCAA GGTCATGGAG AACCAGTCAG AGAGATACAT
 601 CCGATGTCAA GGAACTAATT CCAGAGTTCT ATTACGTACC AGAGATGTTT
 651 GTCAACAGCA ATGGGTACCA TCTTGGAGTG AGGGAGGACG AAGTGGTGGT
 701 TAATGATGTG GACCTGCCCC CCTGGGCCAA GAAGCCAGAA GACTTTGTGC
 751 GGATCAACAG GATGGCCCTG GAAAGTGAAT TTGTTTCTTG CCAACTCCAT
 801 CAATGGATTG ACCTTATATT TGGCTACAAA CAGCGAGGGC CAGAGGCAGT
 851 CCGTGCTCTC AATGTTTTCC ACTACTTGAC CTACGAAGGC TCTGTAAACC
 901 TGGACAGCAT CACAGACCCT GTGCTCCGGG AGGCCATGGT TGCACAGATA
 951 CAGAACTTTG CCCAGACGCC ATCTCAGTTG GTCATTGAGC CGCATCCGCC
1001 TAGGACTTCA GCCATGCATC TGTGTTCCCT TCCACAGAGC CCACTCATGT
1051 TCAAAGATCA GATGCAGCAG GATGTGATCA TGGTGCTGAA GTTTCCATCC
1101 AATTCTCCTG TGACTCATGT GGCTGCCAAC ACCCTGCCCC ACCTGACCAT
1151 CCCTGCAGTG GTGACAGTGA CCTGCAGCCG ACTGTTTGCA GTGAACAGAT
1201 GGCACAACAC AGTCGGCCTC AGAGGAGCCC CCGGATACTC CTTGGATCAA
1251 GCACACCATC TTCCCATTGA GATGGACCCA TTAATCGCAA ATAACTCTGG
1301 TGTGAACAAG CGGCAGATCA CAGACCTTGT AGACCAGAGC ATCCAGATCA
1351 ATGCCCACTG CTTCGTGGTC ACAGCTGATA ATCGCTACAT CCTCATCTGT
1401 GGGTTTTGGG ATAAAAGTTT CAGAGTTTAC TCGACAGAAA CAGGGAAACT
1451 GACACAGATT GTATTTGGCC ACTGGGATGT TGTCACATGC CTGGCCAGGT
1501 CGGAGTCCTA CATTGGTGGA GACTGCTACA TAGTGTCTGG ATCTCGGGAC
1551 GCCACCTTGC TTCTCTGGTA CTGGAGTGGG CGTCACCACA TCATCGGAGA
1601 CAACCGCAAT AGCAGTGACT ATCCTCGCGC CAGAGCTGTC CTCACAGGCC
1651 ATGACCATGA AGTTGTCTGT GTCTCCGTCT GTGCAGAACT CGGACTCGTT
1701 ATCAGTGGTG CTAAAGAGGG CCCTTGCCTC GTTCATACCA TCACTGGAAA
1751 TCTGCTGAAG GCCCTGGAAG GACCAGAAAA CTGCTTATTT CCACGCCTAA
1801 TTTCGGTATC CAGTGAAGGC CACTGCATCA TATATTATGA GCGAGGACGG
1851 TTTAGCAACT TCAGCATCAA TGGGAAACTT TTGGCTCAAA TGGAGATCAA
1901 TGATTCCACT AGGGCTATTC TCCTGAGCAG CGATGGACAG AACCTGGTGA
1951 CTGGAGGGGA CAATGGTGTG GTGGAGGTCT GGCAGGCCTG TGACTTTAAG
2001 CAGCTGTACA TTTACCCAGG ATGTGATGCT GGCATTAGAG CGATGGATTT
2051 ATCCCATGAC CAAAGGACTC TGATCACTGG CATGGCTTCC GGCAGCATTG
2101 TACTTTTAAT ATAGATTTTA ATCGGTGGCA TTATGAGCAT CAGAACAGTA
2151 CTGAAGAGAA GCAGCAGAAG CCACATTCAA GTGAGAGCAC AAGTGCTTCT
2201 GTGGAAAGGC AGTATCTCTG GTGGGACGCT GGTCCACATC GGCCTCTGCT
2251 TGTACATCCA TCCCACCCAG CAGTCGCCGA ACATCATAGT CGGGAGCCAT
2301 TCACCCTGGT ATTCAGGA CTGAACACCA GCTGCTGTCA AGCAAGCTTA
2351 TATCATGTAA ATTATCTGAA TTAGGAGCCG TTTTGGTAAT TATTTATAT
2401 ATCGCCGTTT ATTGAGAAAA GGTTGTAGGA AGCCTCACAA GAGACTTTTG
2451 ACAATTCTGA GGAACCTTGT GCCCAGTTGT TACAAAGTTT AAGCTTTGAA
     CCTAACTTGC ATCCCATTTC CAGCCTCGGG CTTCACTCGT GCC
```

d) Mouse Lyst2 peptide:

```
   1 SRANRTSVMF NFPDQATVKK VVYSLPRVGV GTSYGLPQAR RISLATPRQL
  51 YKSSNMTQRW QRREISNFEY LMFLNTIAGR TYNDLNQYPV FPWVLTNYES
 101 EELDLTLPGN FRHLSKPKGA LNPKRAVFYA ERYETWEEDQ SPPFHYNTHY
 151 STATSPLSWL VRIEPFTTFF LNANDGKFDH PDRTFSSIAR SWRTSQRDTS
 201 DVKELIPEFY YVPEMFVNSN GYHLGVREDE VVVNDVDLPP WAKKPEDFVR
 251 INRMALESEF VSCQLHQWID LIFGYKQRGP EAVRALNVFH YLTYEGSVNL
 301 DSITDPVLRE AMVAQIQNFA QTPSQLLIEP HPPRTSAMHL CSLPQSPLMF
 351 KDQMQQDVIM VLKFPSNSPV THVAANTLPH LTIPAVVTVT CSRLFAVNRW
 401 HNTVGLRGAP GYSLDQAHHL PIEMDPLIAN NSGVNKRQIT DLVDQSIQIN
 451 AHCFVVTADN RYILICGFWD KSFRVYSTET GKLTQIVFGH WDVVTCLARS
 501 ESYIGGDCYI VSGSRDATLL LWYWSGRHHI IGDNPNSSDY PAPRAVLTGH
 551 DHEVVCVSVC AELGLVISGA KEGPCLVHTI TGNLLKALEG PENCLFPRLI
 601 SVSSEGHCII YYERGRFSNF SINGKLLAQM EINDSTRAIL LSSDGQNLVT
 651 GGDNGVVEVW QACDFKQLYI YPGCDAGIRA MDLSHDQRTL ITGMASGSIV
     LLI*
```

Fig. 11. Partial cDNA (a, c) and predicted amino acid (b, d) sequences of human (a, b) and mouse (c, d) lysosomal trafficking regulator 2, LYST2.

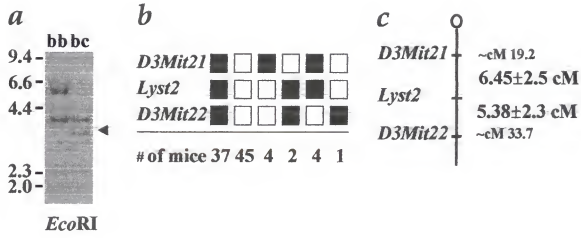

Figure 12. Genetic mapping of *Lyst2* on mouse chromosome 3.

a) Autoradiograph of a Southern blot of mouse genomic DNA, digested with *Eco*RI and hybridized to a *Lyst2* probe. Molecular size standards (in kilobases) are shown to the left. The arrowhead indicates the RFLP, which was used for the genetic mapping and was present in DNA from (C57BL/6J-*bg*^{J} x CAST/EiJ) F_1 **(bc)** but not in homozygous C57BL/6J-*bg*^{J} **(bb)** DNA.

b) Haplotype analysis of *Lyst2*, *D3Mit21* and *D3Mit22* on mouse chromosome 3 in 93 [C57BL/6J-*bg*^{J} X (C57BL/6J-*bg*^{J} x CAST/EiJ) F_1] intersubspecific backcross mice. Black squares represent the homozygous C57BL/6J pattern and white squares, the F_1 pattern. The number of mice of each haplotype is indicated.

c) Composite genetic map of part of mouse chromosome 3, showing the relative position of the *Lyst2* locus. Marker locations are expressed as genetic distance from the centromere (in centimorgans), and intermarker distances (± standard error) are shown.

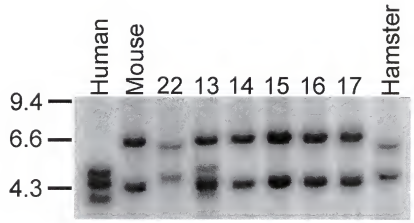

Figure 13. Mapping of *LYST2* on human chromosome 13 by human-rodent somatic cell hybrids. An *Eco*RI monochromosomal somatic cell hybrid Southern blot was hybridized to [α-^{32}P]dCTP-labeled LYST2 probe. Lane 1 contains control human DNA; lane 2 - control mouse DNA; lane 3 - human Chr 22 on hamster background; lanes 4 to 8 - human Chr 13, 14, 15, 16, and 17, respectively, on mouse background; lane 9 – control hamster DNA. Molecular size standards (in kilobases) are shown to the left of the gel.

Expression Analysis

In order to determine the size of the LYST2 transcript(s) and their abundance in different organs, expression analysis was performed using northern blots from various human and mouse tissues.

Transcript size in mouse and human. Hybridization of a *LYST2* probe to northern blots of mouse and human tissues (Fig. 14) revealed that the largest band was ~12-13 kb, very similar to the size of the largest isoform (*LYST*-I) of the Chediak-Higashi syndrome gene (Barbosa *et al.*, 1997). Low levels of additional transcripts of ~6 kb and ~5 kb were visible in mouse brain RNA.

Transcript abundance in mouse. Hybridization to poly(A)$^+$ northern blots from 8 mouse tissues showed that *Lyst2* was abundantly expressed only in brain. Moderate expression was observed in kidney, and weak expression - in heart, lung, skeletal muscle and testis. The gene was not expressed in spleen and liver (Fig. 14a).

Transcript abundance in human. In northern blots of selected human tumor cell lines (Fig. 14b), *LYST2* was moderately expressed in melanoma and colorectal adenocarcinoma, weakly expressed in HeLa, lymphoblastic leukaemia and Burkitt's lymphoma cells and not expressed in promyelocytic and myelogenous leukaemia, or in lung carcinoma lines. In 7 normal human tissues, *LYST2* was weakly expressed in spleen, lymph node, thymus and appendix and not expressed in peripheral blood leukocyte, bone marrow and fetal liver (Fig. 14c).

The relative abundance of *LYST* and *LYST2* mRNA in 50 human tissues was compared by sequential hybridization of corresponding gene-specific probes to a

poly(A)$^+$ RNA dot blot (Fig. 15). The quantity of poly(A)$^+$ RNA loaded on the blot was normalized to 8 housekeeping genes (phospholipase, ribosomal protein S9, tubulin, a highly basic 23 kDa protein, glyceraldehyde-3-phosphate dehydrogenase, hypoxanthine guanine phosphoribosil transferase, ß-actin and ubiquitin).

When a probe specific for the largest mRNA isoform of *LYST* was hybridized to the blot, transcripts were most abundant in immune system organs (adult and fetal thymus, peripheral blood leukocytes, bone marrow) and in several regions of adult brain (Fig. 15). There was moderate to low relative *LYST* expression in most of the other tissues tested. Interestingly, no *LYST* expression was detected in fetal brain, as well as in salivary gland, kidney, lung and fetal heart.

Following hybridization of the same blot with a *LYST2*-specific probe, a different pattern of expression was observed (Fig. 15). *LYST2* transcripts were most abundant in whole brain (adult and fetal) and in 14 specific adult brain regions, as well as in kidney, while there was low to negligible *LYST2* expression in the rest of the tissues.

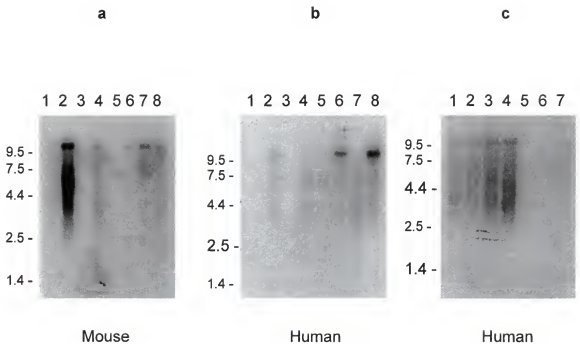

Figure 14. Northern blot analysis of mouse and human LYST2.
Northern blots of 2 μg poly(A)$^+$ RNA from various mouse tissues (**a**), human cancer cell lines (**b**), or human lymphoid tissues (**c**), hybridized to a [α-^{32}P]dCTP-labeled LYST2 probe. Lanes: (**a**) 1-heart, 2-brain, 3-spleen, 4-lung, 5-liver, 6-muscle, 7-kidney, 8-testis; (**b**) 1-promyelocytic leukemia, 2-HeLa, 3-myelogenous leukemia, 4-lymphoblastic leukemia, 5-Burkitt's lymphoma, 6-adenocarcinoma, 7-lung carcinoma, 8-melanoma; (**c**) 1-spleen, 2-lymph node, 3-thymus, 4-appendix, 5-leukocyte, 6-bone marrow, 7-fetal liver. Molecular size standards (in kb) are shown to the left.

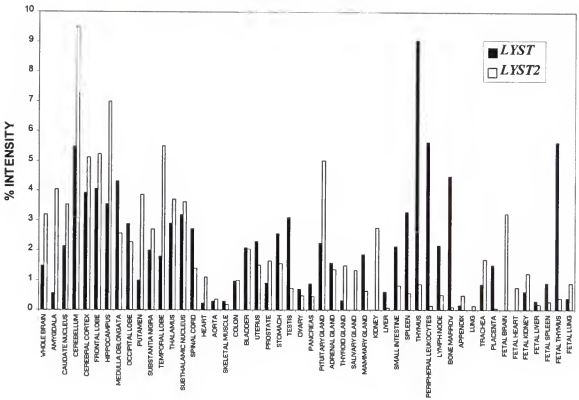

Figure 15. Comparison of the relative abundance of *LYST* and *LYST2* transcripts in 50 human tissues after normalization. Variation of the relative mRNA abundance among tissues was assessed by phosphorimaging, dot quantitation, subtraction of background and calculation of the intensity percentage for each tissue. The relative mRNA abundance between normalized tissues for each probe was expressed as the percentage of the total hybridization signal obtained for each tissue with that probe. The average percentage intensity for housekeeping genes in all tissues was ~2%. No hybridization was evident to control dots, containing yeast total RNA, yeast tRNA, *E. coli* rRNA, *E. coli* DNA, poly r(A), human C_0t1 DNA and human DNA. Exposure time for phosphorimaging was 20 h.

Identification of Proteins that Interact with the CHS Protein and with LYST2, Using a Yeast Two-Hybrid Approach

In an effort to investigate the function of the LYST and LYST2 proteins, two types of screens for protein-protein interactions were performed by the yeast two-hybrid method, as described in MATERIALS AND METHODS. In a "forward" screen, one hybrid consisted of the DNA binding domain of the yeast transcriptional activator Gal4 fused to a "bait" portion of LYST or LYST2. The other hybrid consisted of the Gal4 activation domain fused to "prey" sequences encoded by a mammalian cDNA library. In a "reverse" screen, the "bait" part of LYST or LYST2 was fused to the Gal4 activation domain, and the "prey" sequences of the mammalian cDNA library were fused to the Gal4 DNA binding domain. The activation domain vector contained the *LEU2* gene for selection in yeast deficient in leucine biosynthesis, and the DNA-binding domain cloning vector included the *TRP1* gene for selection in yeast deficient in tryptophane biosynthesis. Each of the vectors was transformed into complementary (a and α) mating types of yeast. Mating was carried out to express both vector constructs within the same yeast cells, thus allowing interaction to occur. Interaction between the bait and prey proteins led to transcriptional activation of reporter genes containing cis-binding elements for Gal4. The reporter genes encoding the indicator protein beta-galactosidase, and metabolic markers for uracil and histidine auxotrophy, were included in specific fashion in one or the other of the yeast strains used in the mating, as described in the MATERIALS AND METHODS chapter. As a result, yeast were selected for expression of both fusion constructs, for successful mating, and for expression of LYST- or LYST2-

interacting proteins (IPs). Yeast clones that contained interacting proteins were grown in individual wells of microtiter plates and the plasmids containing the IP-sequences were isolated and characterized. Numerous tests (for bait self-activation, prey self-activation, non-specific binding, sequencing of all baits, matrix mating, β-galactosidase tests for haploids and for diploids) were performed, as described in the MATERIALS AND METHODS chapter, to determine the specificity of the detected bait:prey interactions. The following interactions, summarized in Table 5, passed all confirmation tests.

In the forward screens, four different molecules were found to interact with the CHS protein. LYST fragment 190-1056 interacted with sequences identical to the 14-3-3 protein. Three independent isolates of 14-3-3 sequence, starting at nucleotides 270, 306, and 330, representing the same region of the molecule (as shown in APPENDIX, Sequence 1), were identified. Another sequence which was found to interact with LYST fragment 3190-4032 was identical to casein kinase II β-subunit. Four independent isolates of CKII-β, starting at nucleotides 1, 4, 31, and 82 (APPENDIX, Sequence 3) and representing the same region of the molecule, were identified. LYST fragment 6586-7449 was found to interact with a sequence identical to human hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs), starting at nucleotide 326 (APPENDIX, Sequence 4). Another identified sequence interacting with LYST fragment 9037-9585 was identical to the EST cg51287.d10 sequence (IP-11, APPENDIX, Sequence 23), starting at nucleotide 1.

In the reverse screens, a total of 22 molecules were identified as LYST- or LYST2-IPs. Ten of those were identical to published proteins and thirteen molecules

were identified as EST sequences (eleven novel and two known ESTs). Identified sequences interacting with the LYST fragment 190-1056 were identical to 14-3-3 protein sequence starting at nucleotides 270, 288 and 354, *i.e.* representing the same domain of 14-3-3 protein identified in the forward screen (APPENDIX, Sequence 1), to B 14-3-3 protein sequence starting at nucleotide 466 (APPENDIX, Sequence 2), to calmodulin sequence starting at nucleotide 474 (APPENDIX, Sequence 5), to estrogen receptor-related protein, starting at nucleotide 756 (as shown in APPENDIX, Sequence 6), and to EST cg50173.d10 sequence (IP-7, APPENDIX, Sequence 19), starting at nucleotide 1. The sequence identical to 14-3-3 protein, starting at nucleotide 270, also interacted with the LYST fragment 4009-4821. LYST fragment 6586-7449 was found to interact with sequences identical to importin β-subunit (starting at nucleotide 1970, APPENDIX, Sequence 7), imogen 38 (starting at nucleotide 492, APPENDIX, Sequence 8), DGS-I (a gene with unknown function, isolated from the DiGeorge syndrome critical region, starting at nucleotide 455, APPENDIX, Sequence 9), EST cg50136.f6 (IP-1, starting at nucleotide 1, APPENDIX, Sequence 3), EST AA010799 (IP-2, starting at nucleotide 1, APPENDIX, Sequence 4), EST cg50136.a7 (IP-4, starting at nucleotide 1, APPENDIX, Sequence 16), EST cg50175.c7 (IP-5, starting at nucleotide 25, APPENDIX, Sequence 17), EST cg50138.g5 (IP-6, starting at nucleotide 1, APPENDIX, Sequence 18), EST cg50175.h7 (IP-8, starting at nucleotide 1, APPENDIX, Sequence 20), EST cg50136.a5.b (IP-10, starting at nucleotide 1, APPENDIX, Sequence 22), and EST cg49432.h3 (IP-12, starting at nucleotide 1, APPENDIX, Sequence 24). The sequence identical to EST cg50138.g5 (IP-6, starting at nucleotide 1) interacted also with LYST fragment 10576-

11611. Identified sequences interacting with LYST fragment 9502-10590 were identical to atrophin-1 (starting at nucleotide 2660, APPENDIX, Sequence 10), embryonic fyn substrate 2 (Efs2, starting at nucleotide 1665, APPENDIX, Sequence 11), similar to EST cg50136.c10 (IP-3, starting at nucleotide 356, APPENDIX, Sequence 5), identical to EST KIAA0192 (IP-9, starting at nucleotide 5092, APPENDIX, Sequence 21), and similar to EST cg50138.e1.b (IP-13, starting at nucleotide 75, APPENDIX, Sequence 25). The sequence identical to EST cg50136.c10 (IP-3) interacted also with LYST fragment 9037-9585. The sequences identical to casein kinase II β-subunit (starting at nucleotides 1 and 58) interacted not only with the LYST fragment 3190-4032 in the forward screen, but also with LYST fragment 10576-11611 in the reverse screen.

In summary, eleven LYST-interacting sequences were found to be identical to published sequences encoding proteins (14-3-3, B14-3-3, Hrs, calmodulin, casein kinase II β-subunit, Efs2, importin β-subunit, estrogen receptor-related protein, imogen 38, atrophin-1, and DGS-I), two LYST-interacting sequences were found to be highly homologous or identical to published EST sequences (AA010799 and KIAA0192, encoding for IP-2 and IP-9, respectively), and eleven LYST-interacting sequences were found to be novel ESTs (cg50136.f6, cg50136.c10, cg50136.a7, cg50175.c7, cg50138.g5, cg50173.d10, cg50175.h7, cg50136.a5.b, cg51287.d10, cg49432.h3.b, and cg50138.e1.b; encoding for IP-1, IP-3, IP-4, IP-5, IP-6, IP-7, IP-8, IP-10, IP-11, IP-12, and IP-13 respectively).

In the reverse screen searching for LYST2-IPs, one interacting sequence was found to be identical to the published sequence encoding human brain factor 2 (HBF-G2,

starting at nucleotide 1638, APPENDIX, Sequence 2) and one to the same region (starting at nucleotide 354, APPENDIX, Sequence 1) of 14-3-3 protein, identified to interact with LYST. The nucleic acid sequences and corresponding amino acid sequences of all LYST- and LYST2-interacting proteins are shown in APPENDIX, Sequences 1-25. A summary of the described protein-protein interactions is shown in Table 5.

Table 5. Protein interactions identified by forward and reverse yeast two-hybrid screens using fragments of LYST and LYST2 as baits

BAIT FRAGMENT AND POSITION, bp*	INTERACTING PROTEINS IDENTIFIED IN	
	FORWARD SCREENS	REVERSE SCREENS
LYST 190-1056	14-3-3 protein	14-3-3 protein; B14-3-3; Calmodulin; IP-7; Estrogen receptor-related protein
LYST 3190-4032	Casein kinase II β-subunit	
LYST 4009-4821		14-3-3 protein
LYST 6586-7449	Hepatocyte growth factor - regulated tyrosine kinase substrate (Hrs)	Importin β-subunit; Imogen 38; DGS-I (DiGeorge syndrome); IP-1,-2,-4,-5,-6,-8,-10,-12
LYST 9037-9585	IP-11	IP-3
LYST 9502-10590		Atrophin-1; Embryonic Fyn substrate 2 (Efs2); IP-3,-9,-13
LYST 10576-11611		Casein kinase II β-subunit; IP-6
LYST2 774-1424		14-3-3 protein; Human brain factor-2 (HBF-G2)

* According to GenBank accession number U67615 for LYST; for LYST2, fragment positions are relative to the beginning of the known sequence

The potential functional significance of the interactions between LYST/LYST2 and the previously published proteins will be discussed in the next chapter. A short description of all ESTs, detected as LYST-IPs, and their similarity to known sequences follows.

IP-1 (EST cg50136.f6, APPENDIX, Sequence 13)

One identified sequence (cg50136.f6, 503 nucleotides), interacting with LYST, was 98% identical to nucleotides 1-128, 90% to nucleotides 339-455, and 71% to nucleotides 201-445 of the human EST AA452346. This EST of 455 nucleotides was initially obtained from a human Soares total fetus (8-9 week) library (Hillier *et al.*, 1997). However, EST AA452346 could be not be used for an extension of EST cg50136.f6 in either direction.

EST cg50136.f6 showed 68% nucleotide identity to mouse T-complex protein Tcp-10 genes (GenBank accession No. M73509, M73506, M22601, M73505). An open reading frame could be translated from nucleotides 3 to 503. The resulting 167 amino acids correspond to a c-terminal or core region of a novel protein, which was designated IP-1. Amino acids 2-126 of IP-1 show 68% homology to the c-terminal region (amino acids 242-365) of mouse Tcp-10 protein (GenBank accession No. X58170)

Therefore, IP-1 represents a novel protein with homology to mouse Tcp-10 protein. Interestingly, another LYST-interactant, IP-10 (see later in this chapter), shows homology to the amino-terminal region of Tcp-10. In addition, LYST-interactant IP-11 has a region homologous to ATP-binding proteins and chaperonin Tcp-1. The nucleotide and amino acid sequences of IP-1 are shown in APPENDIX, Sequence 13.

IP-2 (EST AA010799, APPENDIX, Sequence 14)

Three identical sequences (cg50136.a4, cg50136.d6, cg50136.c9), interacting with LYST, were 96% identical to nucleotides 1-441 of the human EST AA010799. This EST of 495 nucleotides was initially obtained from a human Soares fetal heart (19 weeks) library (Hillier *et al.*, 1995). EST AA010799 could not be further extended by database searches and was used for further analysis.

The 5' end of this EST was similar to the human 50 S ribosomal protein L17. An open reading frame could be translated from nucleotides 3 to 494 of EST AA010799. The resulting 164 amino acids correspond to a c-terminal or core region of a novel protein, which was called IP-2. Amino acids 1-107 of IP-2 show 55% homology to amino acids 5-117 of bacterial ribosomal protein L17 (GenBank accession No. S07223, M26414, L33834, AE00633, D90905).

Therefore, IP-2 represents a novel human protein with homology to ribosomal protein L17. The nucleotide and amino acid sequences of IP-2 are shown in APPENDIX, Sequence 14.

IP-3 (EST cg50136.c10, APPENDIX, Sequence 15)

Two identical sequences (cg50136.c10 and cg50136.g10), interacting with LYST, showed significant homology to published ESTs. An extended expressed sequence of 1198 nucleotides was assembled from nucleotides 1 to 466 of human fetal lung EST W40354 (Hillier *et al.*, 1995), nucleotides 57 to 194 from cg50136.c10 (corresponding to nucleotides 345 to 539 in the extended sequence), nucleotides 1 to 382 from human endothelial cell EST AA186481 (Hillier *et al.*, 1995), corresponding to nucleotides 483 to

863 in the extended sequence, nucleotides 336 to 446 from human fetal EST AA436726 (Hillier *et al.*, 1997), corresponding to nucleotides 529 to 974 in the extended sequence, and nucleotides 199 to 423 from EST N49053 from a human multiple sclerosis library (Hillier *et al.*, 1995), corresponding to nucleotides 777 to 1198 in the extended sequence. The fragment interacting with LYST starts at nucleotide 345. The extended sequence of 1198 nucleotides was used for subsequent analysis.

Nucleotides 213 to 599 of the extended expressed sequence showed 75% identity to nucleotides 3589 to 3975 of rat Olf-1/EBF associated Zn finger protein Roaz (GenBank accession No. U92564). An open reading frame could be translated from nucleotides 1 to 594. These 198 amino acids represent a core or c-terminal region of a novel protein, which was designated IP-3. Amino acids 26 to 197 of IP-3 show 84% homology to amino acids 1023 to 1185 of the rat Zn finger protein Roaz. Weaker homologies (40 to 50%) were seen to different zinc-finger proteins of several species. Interestingly, conserved zinc-finger domains exist in other proteins, *e.g.*, protein kinase C and Hrs, that may be associated with LYST (see next chapter).

Therefore, IP-3 represents a novel human protein with homologies to rat Olf-1/EBF associated Zn finger protein Roaz. The nucleotide and amino acid sequences of IP-3 are shown in APPENDIX, Sequence 15.

IP-4 (EST cg50136.a7, APPENDIX, Sequence16)

One identified prey sequence (cg50136.a7, 493 nucleotides), interacting with LYST, was 91% identical to human EST AA009453, starting at nucleotide 1. This EST of 505 nucleotides was initially obtained from a human Soares fetal heart (19 weeks)

library (Hillier *et al.*, 1997). EST AA009453 could be extended in the 5' direction with nucleotides 57 to 76 of clone cg50136.a7, resulting in a sequence of 524 nucleotides.

An open reading frame could be translated from nucleotides 2 to 244. These 81 amino acids correspond to a carboxy-terminal, proline-rich region of a novel protein, which was called IP-4.

Therefore, IP-4 represents a novel, proline-rich protein without homology to known proteins. The nucleotide and amino acid sequences of IP-4 are shown in APPENDIX, Sequence 16.

IP-5 (EST cg50175.c7, APPENDIX, Sequence 17)

One identified prey sequence (cg50175.c7, 548 nucleotides), interacting with LYST, was 98% identical to human cDNA clone AA403189, starting at nucleotide 24. This EST of 386 nucleotides was initially obtained from a human Soares total fetus (8-9 week) library (Hillier *et al.*, 1997). EST AA403189 was extended with nucleotides 418 to 548 of clone 50175.c7, resulting in a sequence of 517 nucleotides. This sequence could not be extended further by database searches and was used for subsequent analysis.

An open reading frame could be translated from nucleotides 1 to 288. This corresponds to a 96 amino acid protein, which was designated IP-5. Amino acids 10 to 84 of IP-5 were found to be 48% homologous to *Bacillus subtilis* ATP-dependent protease La (GenBank accession No. X76424 and P37945), to *Bacillus subtilis* ATP-dependent Lon protease (GenBank accession No. Z75208), and to *Bacillus subtilis* endopeptidase La (GenBank accession No. I40421).

Therefore, IP-5 represents a novel protein with homology to ATP-dependent proteases Lon/La. The nucleotide and amino acid sequences of IP-5 are shown in APPENDIX, Sequence 17.

IP-6 (EST cg50138.g5, APPENDIX, Sequence 18)

Two identified sequences (cg50138.g5, 498 nucleotides, and the identical cg50138.f2), interacting with LYST, showed significant homology to published expressed sequences. Nucleotides 222 to 492 of cg50138.g5 were 98% identical to nucleotides 1 to 270 of EST AA460131 (588 nucleotides; Hillier *et al.*, 1997). This sequence was obtained from a human total fetus (9 weeks) cDNA library. A longer sequence was assembled from nucleotides 1-449 of cg50138.g5 and nucleotides 228-588 of AA460131, and the extended sequence of 753 nucleotides was used for subsequent analysis.

The extended sequence showed homology to several viral nonstructural proteins: nucleotides 2 to 487 showed 83% identity to nucleotides 170 to 653 of the nonstructural protein Ns2-3 of a border disease virus strain (pestivirus type 3) (GenBank accession No. U43603), nucleotides 42 to 405 showed 90% identity to nucleotides 17 to 378 of proteins p54, p80 and p125 of the bovine viral diarrhea virus (GenBank accession No. Z54331). Nucleotides 311 to 753 showed 58% identity to nucleotides 3025 to 3464 of the human growth hormone and chorionic somatomammotropin genes (GenBank accession No. J03071).

An open reading frame could be translated from nucleotides 1 to 582. These 194 amino acids represent a core or c-terminal region of a novel protein, which was called IP-

6. Amino acids 6 to 144 of IP-6 show 93% homology to amino acids 64 to 198 of the viral (pestivirus type 3) nonstructural protein Ns2-3 (GenBank accession No. U43603). Amino acids 31 to 135 are 98% homologous to amino acids 22 to 126 of the nonstructural protein p125 of the bovine viral diarrhea virus (GenBank accession No. Z54332). BLOCKS analysis searching for similarities to known protein families showed some homology to 14-3-3 proteins (amino acids 17 to 68; identities are shown in capital letters): Efl SKlQD dLKea mntmm CSRcQ GkhRr Femdr Epksa RycAE cnrlh pAE.

Therefore, IP-6 represents a novel human protein with homology to viral nonstructural proteins. The nucleotide and amino acid sequences of IP-6 are shown in APPENDIX, Sequence 18.

IP-7 (EST cg50173.d10, APPENDIX, Sequence 19)

One identified sequence (cg50173.d10, 449 nucleotides), interacting with LYST, showed homology to published expressed sequences. Nucleotides 79 to 325 of cg50173.d10 were 91% identical to nucleotides 30 to 283 of the expressed sequence H62553 (293 nucleotides; Hillier *et al.*, 1995). This sequence was obtained from a human male fetal (20 weeks) liver and spleen cDNA library. The expressed sequence cg50175.d10 could not be further extended by database searches and was used for further analysis.

An open reading frame could be translated from nucleotides 17 to 289, starting with a methionine start codon. These 91 amino acids correspond to a novel protein with a calculated molecular weight of 10589,9, which was designated IP-7. Both the nucleotide and the amino acid sequence of IP-7 show some homology to Alu-domains.

Therefore, IP-7 represents a novel human protein without homology to known proteins. The nucleotide and amino acid sequences of IP-7 are shown in APPENDIX, Sequence 19.

IP-8 (EST cg50175.h7, APPENDIX, Sequence 20)

One identified sequence (cg50175.h7), interacting with LYST, showed significant homology to published expressed sequences. Nucleotides 254 to 555 of cg50175.h7 were 96% identical to nucleotides 1 to 330 of the expressed sequence T09146. This sequence was obtained from a human infant brain cDNA library (Adams *et al.*, 1993). Nucleotides 61 to 433 of cg50175.h7 were assembled to nucleotides 181 to 330 of T09146, resulting in an extended sequence of 523 nucleotides. This sequence could not be further extended by database searches and was used for further analysis.

Nucleotides 2 to 523 of the assembled expressed sequence were 89% homologous to the rat mRNA for norbin (GenBank accession No. AB006461; Shinozaki *et al.*, 1997). Norbin is a novel brain gene, induced by treatment of tetraethylammonium in rat hippocampal slice and accompanied with neurite-outgrowth in neuro 2a cells. The neurite-outgrowth-related protein norbin may play a role in neural plasticity because of the formation of new synapses.

An open reading frame could be translated from nucleotides 2 to 523. These 174 amino acids correspond to a c-terminal or core region of a novel protein, which was called IP-8 and which was 98% identical to amino acids 486 to 659 of rat norbin brain protein (729 amino acids).

Therefore, IP-8 represents a novel protein that appears to be the human homolog of the rat brain gene norbin. The nucleotide and amino acid sequences of IP-8 are shown in APPENDIX, Sequence 20.

IP-9 (EST KIAA0192, APPENDIX, Sequence 21)

One identified prey sequence of 499 nucleotides interacting with LYST was 96% identical to human mRNA clone D83783 for the KIAA0192 gene, starting at nucleotide 5092. The mRNA for KIAA0192 was initially obtained from a human male myeloblast cell line (Nagase *et al.*, 1996).

An open reading frame without a methionine start codon could be translated from nucleotides 1 to 6372. This sequence represents the carboxy-terminal part of a protein which was designated IP-9.

Since no homologies to known proteins could be identified, it appears that IP-9, encoded by the human mRNA for the KIAA0192 gene, represents a novel protein that interacts with LYST. The nucleotide and amino acid sequences of IP-9 are shown in APPENDIX, Sequence 21.

IP-10 (EST cg50136.a5.b, APPENDIX, Sequence 22)

Two identified sequences (cg50136a5 and cg50136c4), interacting with LYST, showed no significant homology to published sequences. Nucleotides 62-309 were 57% homologous to nucleotides 61 to 310 of the human immunodeficiency virus envelope protein (GenBank accession No. L08337). The interacting sequence of 451 nucleotides was used for further analysis.

An open reading frame could be translated from nucleotides 2 to 451. These 150 amino acids correspond to a c-terminal or core region of a novel protein, which was called IP-10. Amino acids 59 to 139 of IP-10 showed 59% homology to amino acids 29-109 of the human T-complex protein TCP-10 (GenBank accession No. U03399). Other homologies (~50%) were found to several cytoskeletal proteins (*e.g.* myosin I isoform, kinesin precursor, cytokeratin, caldesmon, NUF1, tropomyosin, neurofilament protein, kinesin-like protein KIF1, myosin like protein MLP1, troponin T) and to proteins involved in vesicular transport (intracellular protein transport protein, nuclear fusion protein Bik1, synaptonemal complex protein). BLOCKS analysis showed homologies to postsynaptic proteins, gas vesicles protein, clusterin proteins, clathrin light chain proteins, tropomyosin protein and intermediate filament protein.

Therefore, IP-10 represents a novel protein with homologies to the amino-terminal region of Tcp-10 and with homologies to cytoskeletal and vesicular transport proteins. It is interesting that the LYST-interactant IP-1 shows homology to the c-terminal region of Tcp-10. Furthermore, LYST-interactant IP-11 has homologous domains to ATP-binding proteins and chaperonin Tcp-1. The nucleotide and amino acid sequences of IP-10 are shown in APPENDIX, Sequence 22.

IP-11 (EST cg51287.d10, APPENDIX, Sequence 23)

One identified sequence (cg51287.d10), interacting with LYST, was 90% identical to nucleotides 1-325 of the human EST AA310287. This EST of 470 nucleotides was initially obtained from a human Jurkat T-cells library (Adams *et al.*, 1995). The 5' end of EST AA310287 could be extended with nucleotides 154 to 180 of

EST cg51287.d10, resulting in a sequence of 497 nucleotides, which was used for further analysis.

An open reading frame could be translated from nucleotides 3 to 497. These 165 amino acids correspond to a c-terminal or core region of a novel protein, which was designated IP-11. No significant homologies could be found using the BLASTN search program. Using the BLOCKS analysis program, homologies to ATP-binding proteins and chaperonins TCP-1 proteins could be found.

Therefore, IP-11 represents a novel protein with homologies to ATP-binding proteins and TCP-1. The nucleotide and amino acid sequences of IP-11 are shown in APPENDIX, Sequence 23.

IP-12 (EST cg49432.h3.b, APPENDIX, Sequence 24)

Two identified sequences (cg50136c6 and cg50136d2), interacting with LYST, showed no significant homologies to published sequences. The interacting sequence of 402 nucleotides was used for further analysis.

Nucleotides 27-266 of EST cg49432.h3.b were 70% homologous to nucleotides 1115 to 1348 of the human RNA for the cellular oncogene c-fes (GenBank accession No. X52192).

An open reading frame could be translated from nucleotides 2 to 400. These 133 amino acids correspond to a c-terminal or core region of a novel protein, which was called IP-12. Only weak homologies to known proteins could be found, especially to proline-rich regions of different proteins. For example, the proline rich region of IP-12 consisting of amino acids 57 to 89 was 69% homologous to the TGF-β binding protein 3,

amino acids 57 to 75 were 66% homologous to a region in human semaphorin IV, 64% homologous to StpC139 (*Saimiriine herpesvirus 2*), 63% homologous to a collagen-like protein and 55% homologous to homeobox protein HOX-A4. Although the functions of proline-rich regions are often unclear, they could be involved in protein-protein interactions. Some transcription factors contain a proline-rich sequence in the transcriptional activation domain, which has no DNA-binding activity but is essential for activating transcription (Mitchell and Tjian, 1989). Amino acids 14 to 67 of IP-12 showed 53% homology to ORF homolog membrane peroxisomal 70 kDa protein P34230 (ALDP), involved in adrenoleukodystrophy.

Therefore, IP-12 represents a novel protein with a proline-rich region. Other LYST interactants, IP-4 and Hrs, have also proline-rich regions. The nucleotide and amino acid sequences of IP-12 are shown in APPENDIX, Sequence 24.

IP-13 (EST cg50175.c11, APPENDIX, Sequence 25)

One identified sequence (cg50175.c11, 481 nucleotides), interacting with LYST, showed significant homology to published expressed sequences. Nucleotides 1 to 355 of cg50175.c11 were 96% identical to nucleotides 74 to 428 of the expressed sequence H51347 (428 nucleotides; Hillier *et al.*, 1995), obtained from a mouse brain cDNA library. The expressed sequence cg50175.c11 was extended in the 5' direction with nucleotides 1-73 of H51347. This extended sequence showed 96% identity to an expressed sequence Hs21767 of the same length. The extended sequence of 554 nucleotides was used for further analysis.

The extended expressed sequence of 554 nucleotides showed 71% identity to nucleotides 463 to 982 of *Xenopus laevis* elav-type ribonucleoprotein etr-1 (GenBank accession No. U16800). An open reading frame could be translated from nucleotides 3 to 554. These 184 amino acids represent the core or c-terminus of a novel protein, which was designated IP-13. The amino acid sequence of IP-13 shows homologies to ribonucleoproteins of different species, including 67% homology to *Xenopus laevis* elav-type ribonucleoprotein etr-1 (GenBank accession No. U16800) and 54% homology to the human etr-3 protein (GenBank accession No. U69546).

Therefore, IP-13 represents a novel human protein with homology to ribonucleoproteins. The nucleotide and amino acid sequences of IP-13 are shown in APPENDIX, Sequence 25.

DISCUSSION

Since the predicted mouse and human CHS proteins were novel and unlike any of the molecules previously implicated in vesicular transport, their identification did not prove immediately helpful in establishing the precise mechanism whereby CHS dysregulates protein transport and lysosomal trafficking. Therefore, the main goal of this dissertation was to characterize the Chediak-Higashi gene and its products in more detail by using functional genomics approaches:

- identifying alternatively spliced *Lyst* mRNA isoforms and therefore resolving the discrepancy between the initially reported *Lyst* sequences
- identification of mutations in Chediak-Higashi patients and beige alleles
- identification of proteins that interact with LYST

Identification and Characterization of *Lyst* mRNA Isoforms

In order to investigate the expression of the CHS gene and to explain the discrepancy between the previously published *Lyst* sequences, extensive northern and RT-PCR analyses were performed. We demonstrated that each of the previously reported *bg* gene sequences (Barbosa *et al.*, 1996; Perou *et al.*, 1996b) is derived from a single gene with alternatively spliced mRNAs (Fig. 8). The previously reported sequences are

derived from non-overlapping parts of two *Lyst* mRNA isoforms with different predicted C-terminal regions. By sequencing RT-PCR products, we have shown that nucleotides 1-4706 of *Lyst* also represent the previously undetermined 5' region of the largest *Lyst* isoform. Alternative splicing at nucleotide 4706 results in *bg* gene isoforms that contain different 3' regions. Splicing from exon σ (containing nucleotide 4706) to exon τ results in an ~12 kb mRNA (*Lyst*-I) that corresponds to the largest band observed on northern blots. Incomplete splicing at nucleotide 4706 results in a 5893 bp cDNA (*Lyst*-II) that contains intron-derived sequence at the 3' end. *Lyst*-II corresponds to a smaller mRNA observed on northern blots (Barbosa *et al.*, 1996). While several other genes generate an alternative C-terminus by incomplete splicing (Myers *et al.*, 1995; Sugimoto *et al.*, 1995; Sygiyama *et al.*, 1996; Zhao and Manlley, 1996; Van De Wetering *et al.*, 1996), the *bg* gene is unique in that the predicted structures of the two C-termini are quite different. The C-terminus of Lyst-I contains a WD-repeat domain that is similar to the β-subunit of heterotrimeric G proteins and which may assume a propeller-like secondary structure (Lambright *et al.*, 1996). This domain is absent in Lyst-II.

Northern blots of human tissues had suggested that transcription of the homologous human gene, *LYST*, had a similar complexity to the mouse. We identified two human ESTs homologous to mouse Lyst and described a mutation in one of these in a Chediak-Higashi patient (Barbosa *et al.*, 1996). Subsequently, another group published the cDNA sequence of the largest *LYST* isoform, and identified mutations in this gene in two additional patients with CHS (Nagle *et al.*, 1996). We have described the identification of a second isoform of the human gene. This mRNA encodes a protein of

1531 amino acids that is homologous to mouse *Lyst*-II. Like the latter, this human mRNA arises from incomplete splicing and retention of a transcribed intron that encodes the C-terminus of the predicted LYST protein. The mouse and human codons unique to this short isoform share 65% amino acid identity. The stop codon, however, is not conserved precisely between the human and mouse short isoforms. While mouse Lyst-II is predicted to contain a C-terminal prenylation motif (CYSP), translation of human short isoform is predicted to terminate 22 codons earlier and to lack this motif.

However, the other predicted structural features of the human short isoform were conserved with mouse. The most notable of these was a region similar in sequence to stathmin (amino acids 376-540) (Barbosa *et al.*, 1996). While the mouse and human CHS genes had an overall amino acid identity of 81%, identity in the stathmin-like domain was 92% and similarity was 99%. Stathmin is a coiled-coil phosphoprotein that regulates microtubule polymerization in a phosphorylation-dependent manner, and acts as a relay for intracellular signal transduction (Sobel, 1991; Belmont and Mitchison, 1996; Marklund *et al.*, 1996). This region of the *LYST* gene may encode a coiled-coil protein interaction domain and may regulate microtubule-mediated lysosome trafficking. Intriguingly, a defect in microtubule dynamics has been documented previously in CHS (Oliver *et al.*, 1975) and intact microtubules are required for maintenance of lysosomal morphology and trafficking (Matteoni and Kreis, 1987; Swanson *et al.*, 1987; Swanson *et al.*, 1992; Oka and Weigel, 1983).

Mutation and expression analyses suggested that the largest Lyst splice variant (Lyst-I) represents the isoform of primary functional significance and that expression of the smaller isoform (Lyst-II) alone cannot compensate for loss of Lyst-I. This hypothesis

was supported by the genetic complementation experiments initiated in collaboration with Dr. Stephen Brandt. These studies revealed that some beige mouse fibroblasts, expressing full-length Lyst-I, demonstrated more peripheral distribution in the cytoplasm and slight reduction of the size of the stained granules. Such changes were not visible in cells expressing the smaller *Lyst*-II isoform. However, since the effects of the complementation were not very prominent and were not observed in all cells expressing *Lyst*-I, the significance of the described genetic complementation is questionable.

Comparison of the relative abundance of *LYST* gene transcripts in human tissues at different developmental stages revealed an overlapping but distinct pattern of expression (Fig. 9). A quantitative estimate of the relative expression of the smaller mRNA isoforms was generated by comparing the relative hybridization intensity obtained with a probe specific for the large isoform with that obtained with a probe that hybridizes to all *LYST* transcripts. Large isoform transcripts predominated in thymus, fetal thymus, spleen and brain (with the exception of amygdala, occipital lobe, putamen and pituitary gland). Large and small mRNA transcripts were abundant in the latter brain tissues, peripheral blood leukocytes and bone marrow. However, in several tissues, only the small isoforms were expressed, *e.g.* in fetal heart, salivary gland, kidney, lung, and fetal brain. The developmental pattern of mRNA isoform expression in brain was particularly interesting, since only the small isoforms were expressed in fetal brain, whereas the largest isoform predominated in many regions of the adult brain.

Identification of Mutations in Patients with Chediak-Higashi Syndrome and in Beige Mice

Mutation analysis was performed in order to provide additional evidence that *LYST*, resp. *Lyst* is indeed the gene responsible for the CHS in human and for the beige mutation in mouse, as well as to determine whether a genotype-phenotype correlation could be established between different genetic defects and specific disease manifestations. Novel mutations were identified within the coding domain of *LYST* in several CHS patients (Table 4 and Fig.3). The genetic lesions in three patients (370, 372, 373) were C→T substitutions that resulted in premature termination. Another patient was heterozygous for a dinucleotide deletion that results in a frameshift and premature termination. Interestingly, all *bg* and *LYST* mutations identified to date are predicted to result in the production of either truncated or absent proteins, suggesting that missense mutations may not be likely to cause CHS. Unlike Fanconi anemia, type C (Yamashita *et al.*, 1996) there does not appear to be a correlation between the length of the truncated proteins (which may or may not be stable) with clinical features or disease severity in CHS patients, given that a CHS patient with mild manifestations has a truncating mutation early in the protein. However, until the other mutant allele in the compound heterozygote patients is identified, and the exact effects of each mutation at the protein level are characterized, such correlation is imprecise.

Mutation and expression analyses were also useful in addressing the important question about the biological relevance of the 2 main *LYST* transcripts in regulating lysosomal trafficking and in preventing Chediak-Higashi syndrome. The distribution of

expression of mRNA isoforms in human tissues was consistent with the pattern of clinical features observed in CHS patients: transcripts were most abundant in peripheral blood leukocytes, bone marrow, thymus, lymph node, spleen and brain, and common clinical features among CHS patients are immune deficiency, platelet storage pool deficiency and neurologic manifestations. Northern blots suggested that an ~ 13 kb mRNA (corresponding to the largest isoform) is severely reduced in abundance in two CHS patients (Fig. 3). A 4.4-kb band (corresponding to a smaller isoform), however, was present in normal abundance in mRNA from these patients. These results suggest that in some patients, Chediak-Higashi syndrome results from loss of the protein encoded by the largest isoform. Almost all of the mutations identified in CHS patients to date are within the region of the coding domain that is common to all isoforms (Barbosa *et al.*, 1996; Nagle *et al.*, 1996). However, in one CHS patient described by Karim *et al.*, 1997, there was a frameshift (at codon 3197) leading to premature termination in a region of the molecule that belongs only to Lyst-I. In addition, the mouse bg^{kJ} mutation results in the generation of a premature stop codon in *Lyst*-I that is unlikely to affect *Lyst*-II mRNA processing (Perou *et al.*, 1996b). Together, these results suggest that defects in the full-length mRNA alone can elicit Chediak-Higashi syndrome and that expression of the smaller isoform alone cannot compensate for loss of the largest isoform. Although we were not able to establish a more precise genotype-phenotype correlation between different genetic defects and specific disease manifestations, by identifying mutations in several CHS patients and *bg* alleles we did confirm that *LYST/Lyst* is indeed the gene responsible for the Chediak-Higashi syndrome.

Identification and Characterization of *Lyst2*, a Brain-Specific Member of the Chediak-Higashi Syndrome Gene Family

In an effort to identify genes similar to *LYST*, that may play roles in the regulation of vesicular trafficking, we have isolated a novel gene (Fig. 11) in human (*LYST2*) and mouse (*Lyst2*), that exhibits significant sequence identity, similar putative protein domains, and a comparable mRNA size (~13 kb) to the Chediak-Higashi syndrome cDNA (*LYST*). The segment of greatest sequence similarity between LYST and LYST2 (>50% predicted amino acid identity) is located within the BEACH domain of LYST, a region exhibiting sequence conservation to ORFs from *S. cerevisiae* and *C.elegans,* and the human cell division control protein 4-related protein (Nagle *et al.*, 1996). Although the C-terminal residues of LYST2 are not well conserved with LYST, they do have a similar predicted secondary structure. This region of LYST contains WD repeats and is predicted to assume a propeller secondary structure, similar to the ß-subunit of heterotrimeric G proteins. Furthermore, the stop codons of mouse *Lyst* and human *LYST2* occur approximately the same distance from the matching region. There is also high degree of homology (96.2 % identity in 557 amino acids) between human LYST2 and mouse Lyst2 proteins. Based upon these observations, LYST2 is suggested to be a novel member of the LYST gene family. Since the 5' end of *LYST2* has not been identified yet, the level of homology between LYST and LYST2 may change as new sequence is obtained.

While *Lyst* maps on mouse Chr 13 (Kingsmore *et al.*, 1996a; Kingsmore *et al.*, 1996b), *Lyst2* was mapped by linkage analysis between *D3Mit21* and *D3Mit22* on mouse

Chr 3 (Fig. 12). The only known mouse mutation located in this chromosomal region is blebs (*my*) at ~ cM 29.9. Blebs mice exhibit some abnormalities (*e.g.,* pseudencephaly, midcerebral lesions, renal agenesis and hydronephrosis) that correspond to the organs with most abundant *Lyst2* expression (brain and kidney). However, *Lyst2* is not a very likely candidate for the *my* gene because blebs mice have numerous defects (split sternum, preaxial polydactyly and syndactyly of the middle digits; eye, skin and hair abnormalities; ectopia viscerum) in organs, other than brain and kidney. In addition, *my* maps very close to the distal microsatellite marker (*D3Mit22* at ~ cM 33.7) used for the genetic mapping of *Lyst2*.

In human, *LYST* was shown to map on Chr 1 (Fukai *et al.*, 1996; Barrat *et al.*, 1996), while *LYST2* was mapped in a somatic cell hybrid panel to human Chr 13 (Fig. 13). Comparative analysis of the genome organization of human and mouse (Yui, 1997) did not reveal any conserved region between mouse chromosome 3 and human chromosome 13. Therefore, the determination of the precise localization of *LYST2* on human chromosome 13 may potentially lead to the discovery of a new syntenic group between these autosomes.

Northern blot analysis of human and mouse tissues (Fig. 14) suggested that the transcription of *LYST2* had a similar complexity to that of *LYST*, with several, alternatively spliced isoforms. The size of the largest and most abundant *LYST2* mRNA isoform is ~13 kb, which is very similar to the size of the largest *LYST* isoform. There is also evidence that smaller *LYST2* transcripts (resembling in size those of *LYST*) exist, at least in mouse brain, but possibly in other mouse and human tissues. In contrast, there were considerable differences between *Lyst* and *Lyst2* in their tissue distribution of

expression. While *Lyst* is expressed relatively uniformly in different normal mouse tissues (Barbosa *et al.*, 1996), *Lyst2* exhibited abundant expression only in brain. In addition, smaller *Lyst* transcripts (~4 kb) were much more abundant (Barbosa *et al.*, 1996) than those of *Lyst2* (Fig 14a).

The expression patterns of *LYST* and *LYST2* in human tumor cell lines were also different (Fig. 14b). While the predominant *LYST* isoform is ~3-4 kb, the only *LYST2* transcript is ~13 kb. Similarly, in normal human immune tissues, the only visible *LYST2* isoform was ~13 kb, while the most abundant *LYST* transcript was ~3-4 kb. In both normal and tumor human tissues the expression of *LYST* was more uniform than that of *LYST2*.

Analysis of the relative abundance of *LYST* and *LYST2* transcripts in 50 normal human tissues demonstrated that there are considerable differences in the expression patterns (Fig. 15). *LYST* exhibited highest levels of expression in tissues of the immune system (*e.g.*, adult and fetal thymus, peripheral blood leukocytes, bone marrow, spleen) and in some adult brain regions (*e.g.*, cerebellum, medulla oblongata, frontal lobe, cerebral cortex, hippocampus). In contrast, *LYST2* was expressed primarily in whole brain (adult and fetal), in most brain regions and in kidney, and demonstrated low relative expression in the immune system organs.

Notably, there were some tissues, where *LYST2* was expressed, but no *LYST* transcripts were detected, *e.g.,* kidney, fetal brain, fetal heart, salivary gland and lung. However, there were also organs with high relative abundance of *LYST* mRNA and low relative expression of *LYST2*, mainly tissues of the immune system: thymus (adult and

fetal), peripheral blood leukocyte, bone marrow, spleen, lymph node, as well as in placenta and testis.

It is unlikely that *LYST2* probes cross-hybridize to *LYST* sequences because of the unambiguous genetic mapping and because of the observed expression differences between the two genes.

In general, *LYST* demonstrated more uniform pattern of relative expression than *LYST2*, which was expressed in high levels only in brain and to a lesser extent in kidney.

We identified a novel gene in human (*LYST2)* and mouse (*Lyst2*) that appears to be a relative of the Chediak-Higashi syndrome gene, based on sequence similarity, predicted protein structure and on similar transcript size. Comparison of the relative abundance of *LYST2* and *LYST* mRNAs suggests that *LYST2* is expressed abundantly only in brain and therefore it may represent a brain-specific member of the Chediak-Higashi syndrome gene family. These results suggest that these two genes may perform similar functions in different tissues: while *LYST* is expected to participate in the regulation of lysosomal trafficking and degranulation primarily in immune tissues, *LYST2* may play a similar role in cells of the nervous system. This hypothesis is supported by the yeast two-hybrid identification of a protein from the 14-3-3- family that interacts with both LYST and LYST2 (see below). 14-3-3 proteins have been shown to be involved in the regulation of Ca^{2+}-dependent exocytosis of secretory vesicles (Chamberlain *et al.*, 1995; Roth *et al.*, 1995), as well as in signal transduction (Fu *et al.*, 1994; Reuther *et al.*, 1994).

Identification of Proteins that Interact with the CHS Protein, Using a Yeast Two-Hybrid Approach

In a further effort to elucidate the function of the CHS protein, a modified yeast two-hybrid system was used for identification of proteins that interact with LYST and LYST2. Fourteen overlapping cDNA fragments, covering the entire coding domain of *LYST* and 4 fragments covering the known *LYST2* sequence, were PCR amplified and cloned in binding- and activation-domain vectors containing different selectable markers. The resulting constructs were used as baits in forward and reverse screens of a human fetal brain cDNA library. Numerous tests (for bait self-activation, prey self-activation, non-specific binding, sequencing of all baits, matrix mating, β-galactosidase tests for haploids and diploids) were performed to determine the specificity of the detected bait:prey interactions.

A total of 24 interacting sequences (11 identical to published proteins, 2 highly homologous to published ESTs and 11 novel) passed all confirmation tests (Table 5). Interestingly, several of the 11 previously published proteins participate in the regulation of vesicular transport:

1. 14-3-3 proteins – a family of acidic proteins with molecular masses of ~ 30 kDa that are highly conserved over a wide range of eukaryotic organisms, including plants, yeast, insects, amphibians and mammals. Highest levels of 14-3-3 proteins were found in brain, but all animal and plant tissues examined to date contain several isoforms (Aitken *et al.*, 1992, 1995). Several important functions of 14-3-3 proteins have been described:

- Activation of tyrosine and tryptophane hydroxylases, the rate-limiting enzymes involved in neurotransmitter (catecholamine and serotonin) biosynthesis (Ishimura *et al.*, 1987). Although tyrosine hydroxylase was activated by PKC and c-AMP-dependent protein kinase phosphorylation, the 14-3-3 protein was necessary for activation after phosphorylation by Ca^{2+}/calmodulin-dependent (CAM) kinase II.

- Regulation of protein kinase C (PKC) activity. There have been conflicting reports over the role of 14-3-3 as regulators of PKC. Toker *et al.*, 1990, 1992 found that 14-3-3 inhibited PKC, while Isobe *et al.*, 1992 showed that 14-3-3 activated PKC approximately 2-fold. Nevertheless, interaction of 14-3-3 with PKC is established and it may be that the interaction of different 14-3-3 isoforms with PKC is responsible for the different results obtained. In addition, 14-3-3 proteins are phosphorylated by PKC itself, but not by a wide range of other protein kinases.

- Interaction and activation of Raf protein kinase, a convergent point for different signals that regulate growth factor and hormone effects on differentiation and proliferation (Fu *et al.*, 1994). Interactions of 14-3-3 with other signal transduction molecules have been described (Reuther *et al.*, 1994; Burbelo and Hall, 1995; Morrison, 1994), suggesting a role of the 14-3-3 family as modulators of signaling proteins.

- Stimulation of Ca^{2+}-dependent exocytosis in adrenal chromaffin cells (Burgoyne *et al.*, 1993). This action of 14-3-3 is potentiated by PKC, which suggests a role for 14-3-3 in PKC-mediated control of exocytosis. These results are supported by the observation of membrane-bound 14-3-3 proteins, found at high concentration on the membrane of synaptic vesicles (Martin *et al.*, 1994). In addition, 14-3-3 proteins have been shown to selectively stimulate the release of arachidonic acid from choline and ethanolamine

glycerophospholipids (Zupan *et al.*, 1992). Arachidonic acid stimulates secretion from many cell types, which may be relevant to the role of 14-3-3 in exocytosis. In plants, 14-3-3-mediated increase in exocytosis has been proposed as a major mechanism involved in defense responses (Aitken *et al.*, 1992). By using stage-specific assays for exocytosis in chromaffin cells, it has been shown that 14-3-3 proteins, similarly to α-SNAP, enhanced exocytosis by acting in an early, priming stage (Chamberlain *et al.*, 1995). However, the effects of 14-3-3 and α-SNAP showed different Ca^{2+}-dependencies and time courses, suggesting that they had distinct actions in priming. The time courses of action of 14-3-3 and α-SNAP would be consistent with 14-3-3 proteins activating a process which must occur only once for increased secretion, whereas α-SNAP may be needed continuously during secretion to facilitate multiple rounds of vesicle recruitment, docking and fusion. It has been demonstrated that the stimulation of catecholamine release by 14-3-3 proteins was abolished by prior incubation with the actin filament stabilizing drug phalloidin (Roth and Burgoyne, 1995). The cortical actin network was disassembled and actin was reorganized into intracellular foci following treatment with 14-3-3 proteins. These data suggest that 14-3-3 proteins enhance catecholamine release by reorganization of the cortical actin barrier to allow increased availability of secretory vesicles for exocytosis. It has been shown that other stimuli that activate exocytosis lead to actin reorganization in other cell types, e.g. in mast cells (Koffer *et al.*, 1990) and parotid acinar cells (Perrin *et al.*, 1992). Roth *et al.*, 1994, reported that 14-3-3 is associated with granule membranes and that it can bind phospholipid vesicles and induce their aggregation. Since membrane fusion in exocytosis involves proteins able to modify or reorganize lipids within the

plasma and granule membranes, the ability of 14-3-3 to bind phospholipids and to aggregate phospholipid vesicles suggests that 14-3-3 proteins could play a role in membrane events during exocytosis.

The functions of 14-3-3 proteins in vesicular transport are in close agreement with the expected roles of the CHS protein. As described in the INTRODUCTION of this dissertation, there are numerous examples of impaired vesicle trafficking in CHS – melanosomes are affected in melanocytes, dense granules in platelets, lysosomes and secretory vesicles in various immune cells. These defects indicate that LYST may participate in the regulation of intracellular vesicular transport and in exocytosis, where 14-3-3 proteins have been shown to perform important functions. These observations support the detected association between the CHS protein and 14-3-3. Similarly to the modulation of protein kinase C by 14-3-3, a rapid down-regulation of membrane-bound PKC activity following short-term activation has been described in beige macrophages, NK cells and polymorphonuclear leukocytes (Ito *et al.*, 1988, 1989). In addition, a concavalin A (Con A)-induced, PKC-dependent cap formation was significantly increased in beige PMNs compared to that in control cells (Sato *et al.*, 1990). The decline in PKC activity and the abnormally enhanced cap formation were both abolished by pretreatment of cells with leupeptin, an inhibitor of calpain (a Ca^{2+}-dependent thiol proteinase that hydrolyses PKC). Furthermore, the treatment with leupeptin improved the impaired NK cell activity from *bg/bg* mice, whereas the NK cells from *bg/+* and *+/+* were not affected. These findings suggest that in CHS there is a dysregulation of protein kinase C activity that may alter other signal transduction pathways and cause functional cellular defects. The involvement of 14-3-3 and the CHS protein in both vesicular

transport and signal transduction indicate that these two molecules function in common pathways and supports the LYST:14-3-3 interaction detected by the yeast two-hybrid system.

2. Casein kinase II (CK2, CKII) – serine/threonine protein kinase that is present in the nucleus and cytoplasm of all eukaryotic cells that have been studied. CK2 is a heterotetramer with a molecular mass of ~ 130 kDa, composed of 2 types of subunits (α and β) and has the general structure $\alpha_2\beta_2$ or $\alpha\alpha'\beta_2$. The α and α' are catalytically active, while the β subunit is inactive by itself but it stimulates the catalytic activity of α 5- to 10-fold, causes stabilization of α and more importantly, can change the interactions with substrates and inhibitors (Allende and Allende, 1995). Casein kinase II is unusual among the protein kinases in that it can use both ATP and GTP as phosphate donors. There are more than 100 different protein substrates that are phosphorylated by CK2, many of them involved in DNA replication, transcription and signal transduction. It has been demonstrated that one of the major brain substrates for CK2 is synaptotagmin, an abundant synaptic vesicle protein that has been implicated in vesicle docking and fusion with the plasma membrane (Davletov *et al.*, 1993). Synaptotagmin binds Ca^{2+} and phospholipids at concentrations that are thought to occur physiologically in the nerve terminal during depolarization, suggesting that it may be a regulatory Ca^{2+}-binding protein involved in synaptic vesicle exocytosis. Synaptotagmin is part of a tetramer with multiple Ca^{2+}-binding sites and represents the only known Ca^{2+}-binding protein on the synaptic vesicle surface and the first membrane-bound putative regulatory Ca^{2+}-binding protein (Davletov *et al.*, 1993). Ca^{2+} causes an increase in the phosphorylation of

synaptotagmin by CK2 and not by a Ca^{2+}-dependent kinase. The mechanism of this activation is unclear, but it could be related to the fact that synaptotagmin is a Ca^{2+}- and phospholipid-binding protein that may undergo Ca^{2+}-dependent conformational changes. Synaptotagmin has been reported to interact with several different plasma membrane proteins, including syntaxin, and may have a role in the docking of synaptic vesicles. It is the only one of the major synaptic vesicle proteins with a putative function in vesicular trafficking that is also found on other secretory vesicles with Ca^{2+}-regulated exocytosis (Perin *et al.*, 1991). This suggests that if synaptotagmin functions in the final steps of vesicle exocytosis, these steps may be similar in different secretory pathways. The potential significance of the phosphorylation of synaptotagmin by casein kinase II is supported by the finding that synaptotagmin is a very efficient substrate (better than casein) for CK2 and not for other kinases (Davletov *et al.*, 1993). The phosphorylation site is present in both synaptotagmins I and II and is evolutionarily conserved in different species. Bennet *et al.*, 1993, demonstrated that casein kinase II not only phosphorylates synaptotagmin, but also interacts with a domain of synaptotagmin that is distinct from the phosphorylation site. The functional significance of the CK2:synaptotagmin association is unclear but it probably does not simply represent an enzyme:substrate interaction and may have a modulatory role in nerve terminal function.

The involvement of casein kinase II in synaptic vesicle trafficking was further supported by the findings of Bennet *et al.*, 1993, and Hirling and Scheller, 1996, that CK2 phosphorylates syntaxin and possibly synaptobrevin, the two main neuronal SNAREs, called also t-SNARE and v-SNARE, respectively. The interaction between syntaxin and synaptobrevin is considered one of the critical steps in vesicle docking and

fusion. In general, protein phosphorylation has been well established as a regulatory mechanism for vesicle transport: the synaptic vesicle protein synapsin by Ca^{2+}/calmodulin-dependent protein kinase II regulates its interaction with both synaptic vesicles and actin filaments (Schiebler *et al.*, 1986; Benfenati *et al.*, 1992). This in turn may regulate the availability of synaptic vesicles for release. Similarly, phosphorylation of n-sec1 by protein kinase C has been shown to regulate the n-sec1:syntaxin binding (Fujita *et al.*, 1996). CK2 is also indirectly involved in the regulation of synaptic vesicle transport by phosphorylating calmodulin and PKC, which are known to play important roles in this process. Casein kinase II could be copurified with clathrin-coated vesicles and it could phosphorylate the β-light chain of clathrin, implicating CK2 in vesicular trafficking in general.

LYST has several sites of potential phosphorylation by casein kinase II. However, the CK2:LYST interaction detected by the yeast two-hybrid system is not one of enzyme and substrate since the β-subunit of CK2, which interacts with LYST, has a regulatory and not a catalytic activity. This does not, however, exclude the possibility that LYST is phosphorylated by the catalytic α-subunit of CK2. A similar example of distinct phosphorylation and interaction between synaptotagmin and casein kinase II was described above. The involvement of CK2 in the regulation of synaptic vesicle exocytosis and vesicular trafficking, where LYST is expected to perform its function, further supports the detected interaction between these 2 molecules by the yeast two-hybrid method.

3. Calmodulin (CaM) – an abundant Ca^{2+}-binding protein found in all eukaryotic cells that have been studied. It is a highly conserved, single polypeptide chain with four high-affinity Ca^{2+} binding sites, and it undergoes a conformational change when it binds Ca^{2+}. CaM functions as a multipurpose intracellular Ca^{2+} receptor, mediating many Ca^{2+}-regulated processes. Calmodulin has no enzyme activity by itself but acts by binding to other proteins and altering their activity. Among the targets regulated by Ca^{2+}/calmodulin are various enzymes and membrane transport proteins. In many cells, for example, Ca^{2+}/calmodulin binds to and activates the plasma membrane Ca^{2+}-ATPase that pumps Ca^{2+} out of the cell. Most effects of Ca^{2+}/calmodulin, however, are more indirect and are mediated by Ca^{2+}/calmodulin-dependent protein kinases (CaM-kinases) (Alberts *et al.*, 1994). These enzymes phosphorylate serines or threonines in proteins and the response of a target cell to an increase in free Ca^{2+} concentration depends on which CaM-kinase-regulated target proteins are present in the cell. The best studied example is CaM-kinase II, which is found in all animal cells but is especially enriched in the nervous system where it is highly concentrated in synapses. When neurons that use catecholamines as their neurotransmitters are activated, the influx of Ca^{2+} through the plasma membrane stimulates the cell to secrete their neurotransmitter. The Ca^{2+} influx also activates CaM-kinase II to phosphorylate, and thereby activate, tyrosine hydroxylases, the same enzymes that are activated by the 14-3-3 proteins, as described above. CaM-kinase II can function as a molecular memory device, switching to an active state when exposed to Ca^{2+}/calmodulin and then remaining active, due to autophosphorylation, even after Ca^{2+} is withdrawn. It has been shown that CaM-kinase II phosphorylates a number of nerve

terminal proteins, including syntaxin, synaptobrevin, SNAP-25, αSNAP, NSF and synaptotagmin, indicating a significant involvement of this kinase in neurotransmitter release (Hirling and Scheller, 1996).

There is convincing evidence that calmodulin is involved in neurotransmitter release and secretion in general. It is present in nerve terminals and is associated with synaptic vesicles (DeLorenzo, 1981), as well as with other secretory vesicles: platelet α-granules, chromaffin vesicles and posterior pituitary secretory granules. With the use of different antagonists, calmodulin has been implicated in secretion in a number of cell systems, including adrenal medulla, endocrine pancreas, anterior pituitary, platelets, mast cells, adrenal cortex, nerve terminals, and leukocytes (Trifaro *et al.*, 1992). Calmodulin is one of the proteins from a crude cytosol fraction that will bind to chromaffin granules in the presence of Ca^{2+}. This binding involves specific proteins on the cytoplasmic surface of the granules. Two major Ca^{2+}-dependent calmodulin binding proteins of the chromaffin granule membrane have been described and one of them has been identified as the integral membrane protein synaptotagmin (p65), which also interacts with and is phosphorylated by casein kinase II, as described above. Although the significance of the granule membrane-CaM binding is unclear, the fact that synaptotagmin is a component of synaptic and secretory vesicles from several types suggests that calmodulin binding to this protein may participate in the control of exocytosis. In agreement with these results, high-affinity binding sites for calmodulin have been found in coated vesicles (Moskowitz *et al.*, 1982). When anti-calmodulin antibodies are introduced into chromaffin cells, secretion is inhibited, indicating involvement of CaM in exocytosis. Similar results have

been obtained with anti-calmodulin antibodies in other systems as well (Trifaro *et al.*, 1992), supporting a role for CaM in secretion. Fusion experiments carried out with plasma membranes and anterior pituitary secretory granules or pancreatic granules have indicated the requirement of calmodulin for vesicle-plasma membrane fusion processes (Watkins and Cooperstein, 1983). Addition of calmodulin to permeabilized chromaffin cells increased Ca^{2+}-dependent norepinephrine release in a dose-dependent and CaM-specific manner. The enhanced norepinephrine release was inhibited by tetanus toxin, indicating that calmodulin plays an important role in exocytosis from chromaffin cells (Matsuda *et al.*, 1994). Similarly, exocytosis of cortical granules caused by an increase of Ca^{2+} content immediately upon fertilization of sea urchin eggs appears to be controlled by Ca^{2+}/calmodulin. Incubation with antibodies against CaM causes a loss of sensitivity to Ca^{2+}. Sensitivity is restored when calmodulin is added in excess to inhibited preparations. Cortical exocytosis is also inhibited by CaM-antagonists added to sea urchin eggs (Kreimer and Khotimchenko, 1995). By using stage-specific assays for ATP-dependent priming and Ca^{2+}-mediated triggering of exocytosis, Chamberlain *et al.*, 1995, demonstrated that calmodulin acted in the ATP-independent triggering step and stimulated exocytosis. An indirect participation of CaM in neurotransmitter release may involve calcineurin, a Ca^{2+}/calmodulin-dependent protein phosphatase that is enriched in neural tissues. Among the serine/threonine protein phosphatases, calcineurin is unique in its requirement for Ca^{2+}/calmodulin activation. A variety of ion channels (both ligand- and voltage-gated) that play major roles in regulating neuronal excitability and transmitter release appear to be negatively regulated by calcineurin. In addition, the

release of glutamate from presynaptic terminals and hormone secretion in pancreatic β cells and pituitary cells also seem to be under negative regulation by calcineurin (Yakel, 1997). A similar indirect role of calmodulin in exocytosis was shown by Burgoyne *et al.*, 1986, through modulation of the activity of caldesmon, a CaM-regulated actin-binding protein. It is likely that increased cytosolic Ca^{2+} following stimulation results in solation of actin filaments crosslinked by caldesmon, and the subsequent binding of caldesmon to the chromaffin granule regulates interactions of the vesicles with the plasma membrane or the cytoskeleton and leads to exocytosis.

The direct and indirect involvement of calmodulin in synaptic vesicle exocytosis and in secretion in general is in close agreement with the expected roles of LYST. Furthermore, the findings that CaM is phosphorylated by casein kinase II (another Lyst-interacting protein) and that through CaM-kinase II it regulates the function of tyrosine hydroxylases (also activated by the LYST-interacting 14-3-3 proteins) and of several nerve terminal proteins (syntaxin, synaptotagmin, synaptobrevin, SNAP-25, etc.), further supports the detected LYST:calmodulin interaction and unites all these LYST-interacting proteins in a common network involved in vesicular trafficking regulation.

4. Hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs) – a 115-kDa protein that was originally discovered due to its rapid phosphorylation in cells treated with hepatocyte growth factor (Komada and Kitamura, 1994). Hrs contains a conserved double zinc finger-like motif, as well as a proline-rich, proline- and glutamine-rich, and coiled-coil regions that may mediate protein-protein interactions. It is expressed in a variety of adult mouse tissues and at different stages of mouse whole embryos (Komada

and Kitamura, 1995). Tyrosine phosphorylation of Hrs was also induced in cells treated with epidermal growth factor and platelet-derived growth factor suggesting that Hrs may play a role in the signal transduction pathways of different growth factors in a variety of cell types.

Bean *et al.*, 1997, demonstrated that Hrs is an ATPase and that it interacts with SNAP-25 (synaptosomal-associated protein 25), a nerve terminal protein that resides on the internal surface of the plasma membrane and interacts with syntaxin (t-SNARE), synaptobrevin (v-SNARE) and αSNAP to form a core complex thought to be an intermediate in a biochemical pathway that is essential for vesicular transport. The binding of Hrs to SNAP-25 is inhibited by Ca^{2+} in the physiological concentration range that supports synaptic transmission. Hrs is expressed abundantly throughout the brain and is present at nerve terminals. Recombinant Hrs inhibits Ca^{2+}-triggered noradrenaline release from permeabilized cells in a dose-dependent and saturable manner. The Hrs-SNAP-25 binding was inhibited by Zn^{2+}, and the previously reported enhancement of neurotransmitter release induced by Zn^{2+} suggests that the inhibition of Hrs-SNAP-25 binding may facilitate transmitter release. These data provide a functional demonstration of the ability of Hrs to regulate the secretory processes through Ca^{2+}- and nucleotide-dependent modulation of vesicle trafficking protein complexes. It is possible that Hrs plays a regulatory role in vesicle docking through its ATPase activity, and in later stages of secretion, possibly membrane fusion, as a result of calcium-regulated interaction with SNAP-25. This hypothesis was supported by the finding that Hrs is peripherally associated with the cytoplasmic surface of early endosomes (Komada *et al.*, 1997). By

subcellular fractionation, Hrs was recovered both in the cytoplasmic and membrane fractions. These results, together with the finding that Hrs is homologous to Vps27p, a molecule essential for protein trafficking through a prevacuolar compartment in yeast, suggest that Hrs is involved in endosomal transport. It is hypothesized that the growth factor-bound receptor tyrosine kinase is internalized and delivered to the early endosome, where it phosphorylates Hrs residing on the cytoplasmic surface of the early endosome. Tyrosine phosphorylation activates Hrs and it stimulates the vesicle transport of the growth factor-receptor complex to the late endosome and lysosome. Therefore, Hrs may be a regulator of several types of vesicular functions (endo- and exocytosis) in a variety of cells.

In addition to its role in the hepatocyte growth factor intracellular signaling pathway, Hrs was found to interact with STAM (signal transducing adaptor molecule), which was shown to be involved in cytokine-mediated intracellular signal transduction (Asao *et al.*, 1997). Tyrosine phosphorylation of Hrs was induced rapidly after stimulation with IL-2 and GM-CSF, as well as HGF. The mutual association sites of Hrs and STAM include highly conserved coiled-coil sequences, suggesting that their association is mediated by the coiled-coil structures. Exogenous introduction of Hrs significantly suppressed DNA synthesis upon stimulation with IL-2 and GM-CSF, while an Hrs mutant deleted of the STAM-binding site lost such suppressive ability, i.e. Hrs-STAM association is a prerequisite for such activity. These results suggest that Hrs counteracts the STAM function which is essential for cell growth signaling mediated by the cytokines.

The involvement of Hrs in both signal transduction and vesicular trafficking correlates well not only with the predicted function of LYST but also with the roles of the other LYST-interacting proteins detected by the yeast two-hybrid method. IL-2, which induces rapid tyrosine phosphorylation of Hrs, also activates the Shc/Grb2/Sos/Ras/Raf1/mitogen-activated protein kinase cascade (Asao *et al.*, 1997). Raf1 has been shown to interact with 14-3-3 proteins (another Lyst-interacting protein), as described above. IL-2 also activates Fyn tyrosine kinase, and the SH2 domain of Fyn has been reported to interact with Raf1 (Morrison, 1995), providing an additional link to 14-3-3 and, therefore, to LYST. This association is supported further by the finding of interaction between 14-3-3 and Bcr, a protein kinase that is involved in the same growth factor signaling Shc/Grb2/Sos/Ras/Raf1/MAP-kinase cascade. On the other hand, it has been demonstrated that Hrs is homologous to Vps27p, a yeast protein localized to the prevacuolar compartment considered to be the counterpart of mammalian endosomes. *VPS27* is one of the genes whose disruption in *S. cerevisiae* dysregulates the traffic of endocytosed proteins and newly synthesized vacuolar proteins to vacuoles (a counterpart of mammalian lysosomes) and causes the accumulation of these proteins in a prevacuolar compartment. The finding that Hrs is localized to the cytoplasmic surface of early endosomes indicates that in mammalian cells Hrs may play a role similar to that of Vps27p in yeast. The involvement of Hrs in Ca^{2+}-regulated secretion and its interaction with SNAP-25, a protein that functions in exocytosis of neurotransmitter vesicles, further supports the LYST:Hrs interaction detected by the yeast two-hybrid system.

The rest of the previously published proteins, identified as LYST-interactants by the yeast two-hybrid screens, are not as obviously involved in vesicular trafficking and its

regulation, as the above described 14-3-3 proteins, casein kinase II, calmodulin and Hrs. However, for most of these proteins there are some functional and/or morphological indications that they may participate in the same pathways as LYST, 14-3-3, CK2, CaM and Hrs.

1. Atrophin 1 – the gene responsible for dentatorubral pallidoluysian atrophy (DRPLA, Smith's disease), a rare autosomal dominant neurodegenerative disorder characterized by cerebellar ataxia, myoclonic epilepsy, choreoathetosis and dementia. The molecular defect in DRPLA is an expansion of a CAG trinucleotide repeat (encoding glutamine) in atrophin 1. The gene is expressed in many mammalian (but not in avian) species, in all brain regions tested, most abundantly in cerebellum. There is no detectable difference in the level or distribution of atrophin 1 expression between normal and DRPLA-brain, indicating that disease pathology may result from altered protein function (Margolis *et al.*, 1996). Immunohistochemical studies demonstrated predominantly cytoplasmic signal with prominent granular and punctate staining within cell bodies and dendrites, suggesting that atrophin 1 may play a role in vesicle transport (Knight *et al.*, 1997; Yazawa *et al.*, 1995). This hypothesis is supported by the finding of 2 atrophin-interacting proteins. Both have WW domains that may interact with PY motifs in atrophin 1. Expression of one of these proteins appears to be restricted to the brain, while the other is expressed in all tissues examined. One of the atrophin-interacting proteins is homologous to members of a family of E3 ubiquitin-protein ligases, including the yeast Rsp5 protein. Interestingly, rsp5 mutants are deficient in endocytosis (Ross *et al.*, 1997), suggesting a possible involvement of atrophin 1 in vesicular transport and supporting the detected LYST:atrophin 1 interaction.

2. Imogen 38 – a novel ~ 38 kDa protein recognized by a T cell clone from a type I diabetic patient. The protein is expressed at variable levels in many tissues and cell lines. There is some discrepancy regarding the subcellular localization of imogen 38 – Arden *et al.*, (1996) showed that the mitochondrion is the major site of the protein (although there was some overlap of the antigen distribution with that of secretory granules), while Roep *et al.*, (1990) determined that imogen 38 was associated with insulin secretory granules. Although the precise function of the protein is not known, its vesicular localization suggests a possible involvement in secretion, indirectly supporting the detected LYST:imogen 38 interaction by the yeast two-hybrid system.

3. Embryonal Fyn substrate 2 (Efs2) – an isoform of the Efs gene, lacking the SH3 domain. Efs was originally discovered through its Fyn-binding capacity. Embryonal Fyn substrate has characteristic regions important in intracellular signal transduction – an SH3 domain, a cluster of putative ligands for SH2 domains and proline-rich sequences with SH3-binding consensus. The protein was named "embryonal Fyn substrate" because its RNA was most abundantly expressed in embryo, although expression was shown in many adult tissues as well (Ishino *et al.*, 1997). The SH3 domain has been shown to mediate specific binding of Sos1 to Grb2 and this important protein-protein interaction links the tyrosine kinase pathway initiated at the membrane surface to the Ras pathway. Efs2, which lacks the SH3 domain, may participate in the regulation of coupling of these signaling proteins. The associations of Efs/Efs2 with the tyrosine kinase Fyn and with the Grb2/Sos1/Ras pathway indicate its possible involvement in the same processes as Hrs and 14-3-3, two other LYST-interacting proteins.

4. Importin β-subunit – a 90-kDa protein that binds the 60-kDa α-subunit, forming a heterodimer, called importin, or karyopherin. The α-subunit recognizes the nuclear localization signal (NLS) of a cargo protein (called karyophile) that is to be transported from the cytoplasm through a nuclear pore into the nucleus, while the β-subunit first docks the α/β/karyophile complex at the cytoplasmic face of the nuclear pore and then mediates the translocation by interacting with the pore proteins. The small GTPase Ran, as well as a cofactor called p10, p15 or NTF-2, are required for nuclear import. RanGTP dissociates α from β, and may promote the translocation of the α/β/karyophile complex across the pore and the release of the complex from the nuclear side of the pore (Laskey *et al.*, 1996). Interestingly, both importin α and β consist largely of repeated domains, the so called arm repeats, which are also present in LYST. These repeats are found in several functionally unrelated proteins, including the founder member of the family, the *Drosophila* armadillo protein (Peifer *et al.*, 1994). Importin α has 8 arm repeats, while importin β has 11 more divergent repeats, although the precise number depends on how the domains are defined. The NLS binds directly to the arm repeat region of the α-subunit. Besides the presence of the arm repeats, the only other common feature of LYST and importin β is that both proteins are involved in protein transport, although importin plays a role in nuclear transport, while LYST is expected to participate in vesicular protein trafficking.

5. Estrogen receptor-related protein (ERRl) – a member of the nuclear hormone receptor superfamily of zinc finger transcription factors. The ERRl differs from the classical nuclear hormone receptors in that it has no known ligand, making it an "orphan

receptor". It has been shown that the transcription activity mediated by the glucocorticoid receptor was strongly suppressed by estrogen receptor-related protein in a cell- and receptor-specific manner (Trapp and Holsboer, 1996). However, ERRl activates estrogen receptor-mediated response of the lactoferrin gene promoter by binding to a regulatory DNA sequence upstream of the estrogen response element (Yang *et al.*, 1996). It has also been demonstrated that ERRl interacts with the estrogen receptor through a direct protein-protein contact, suggesting that ERRl may participate in estrogen stimulation of the lactoferrin gene (and possibly in other estrogen-mediated pathways) and therefore ERRl may play a role in modulating immune responses and cell growth.

6. DiGeorge syndrome-I (DGS-I) – one of the genes from the critical region of DiGeorge syndrome and velo-cardio-facial syndrome on human Chr 22q11. Both conditions could be caused by microdeletions in this minimal region. It has not been demonstrated that DGS-I is responsible for any of the features of DiGeorge syndrome (absence or hypoplasia of the thymus and parathyroid glands, cardiovascular anomalies and mild craniofacial dysmorphia) or velo-cardio-facial syndrome (cleft palate, cardiac anomalies, typical facies and learning disabilities). DGS-I does not show significant homology to any genes with known function and its role is unclear. There are 2 DGS-I isoforms: the small (~ 1.7 kb) is expressed in all tissues tested, while the large (~ 5.2 kb) is expressed in heart and skeletal muscle only (Gong *et al.*, 1996). Since the available information about DGS-I is so limited, it is hard to discuss the interaction between LYST and DGS-I detected by the yeast two-hybrid method.

The rest of the LYST-interacting proteins (total of 13) were novel (some were homologous to ESTs) and they were briefly discussed in the RESULTS chapter of this

dissertation. Notably, one of the interactants (IP-11) is a novel protein with homology to the chaperone TCP-1. Molecular chaperones guide the folding of newly synthesized proteins through a "molten globule" state to the final correct protein conformation. Eukaryotic cells have chaperone proteins of the hsp60 and hsp70 (heat-shock proteins) families. Different family members have specific subcellular distribution and function in different organelles, *e.g.,* mitochondria contain their own hsp60 and hsp70 molecules that are distinct from those that function in the cytosol. A special hsp70 chaperone, called binding protein (BiP), not only helps in protein folding in the endoplasmic reticulum, but also keeps the resident proteins in the ER and thus out of the Golgi apparatus and the rest of the secretory pathway. In CHS cells there is compartmental miss-sorting of proteins such as elastase, β-glucuronidase, and cathepsin G (Holcombe *et al.*, 1994). In addition, the low tyrosinase activity in giant melanosomes results from mislocalization and eventual extracellular secretion of tyrosinase. These observations support the detected LYST:IP-11 interaction since it is possible that the protein miss-sorting, demonstrated in CHS cells, is related to deficient TCP-1-like chaperone function in the absence of LYST.

Using the yeast two-hybrid system, LYST 2 was found to interact with 2 proteins – 14-3-3 and human brain factor 2. The finding of a protein (14-3-3) that interacts with both LYST and LYST2 supports not only the validity of the detected interactions by the yeast two-hybrid system, but also the predicted relationship between LYST and LYST2. Since the involvement of 14-3-3 in vesicular trafficking and signal transduction was already described above, only the role of HBF-G2 will be discussed here.

Human brain factor 2 (HBF-G2) – a member of the fork head family of transcription factors. It is expressed primarily in embryonic brain but also at lower levels in several adult tissues. HBF-G2 shows high homology to human brain factor 1 and to the retroviral oncogene *qin* (Wiese *et al.*, 1995). The cellular homologs of retroviral oncogenes represent important growth-regulatory genes. Although its function is not known, it is hypothesized that HBF-G2 plays a role in cell proliferation during embryonic brain development and possibly in some adult tissues. The findings that both LYST2 and HBF-G2 are expressed most abundantly in brain indirectly supports their interaction detected by the yeast two-hybrid system.

All of the described IPs were identified by screening only a human brain cDNA library. Other LYST- and LYST2-interacting proteins could by isolated if other libraries (*e.g.,* lymphoid) are screened.

An attempt was made to group the previously published LYST- and LYST2-interacting proteins based on their known or predicted function (Figure 16).

Ten of the 11 proteins could be categorized in only 2 overlapping sets: embryonal Fyn substrate 2, estrogen receptor-related protein and human brain factor 2 play roles in signal transduction; atrophin 1, imogen 38 and importin function in protein transport; and 14-3-3 proteins, casein kinase II, calmodulin and Hrs are involved in both signal transduction and protein transport. The only LYST-interacting protein with currently unknown function is DGS-I but in the future, following its characterization, it may also be categorized in one of the above 2 groups.

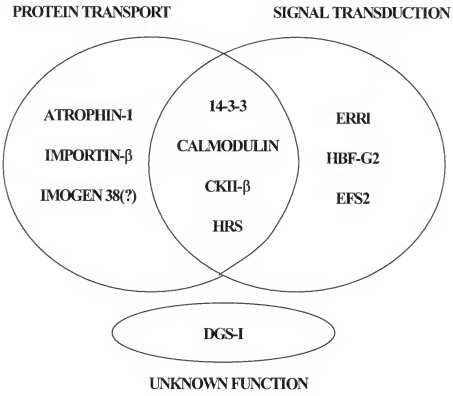

Figure 16. A classification of the previously known LYST- and LYST2-interacting proteins based on their established or predicted function. Almost all of the IPs could be categorized in only 2 overlapping groups – protein transport and signal transduction. Four proteins (14-3-3, casein kinase II, calmodulin and Hrs) belong to both categories and provide a link between the other interacting proteins.

Figure 17 represents a simplified diagram of the interactions between LYST, LYST2 and 4 of their interacting proteins in the context of neurotransmitter vesicle docking/fusion with the presynaptic plasma membrane of a neuron. These 4 LYST-IPs (14-3-3, CK2, calmodulin and Hrs) were selected and included in this diagram because their roles in both signal transduction and vesicular transport are well established and these 4 proteins provide a link between the other LYST-IPs, categorized in either signal transduction or protein transport only. It is shown that LYST, 14-3-3, CK2, calmodulin, Hrs and other proteins known to be involved in vesicular trafficking and signal transduction form a common network of interacting molecules involved in the regulation of these processes.

An independent way of confirming the detected LYST:IP interactions would be to perform coimmunoprecipitation using specific antibodies. While antibodies against some of the LYST-IPs are available, a high-quality LYST-specific antibody has proved difficult to raise, after several attempts by different groups. An antibody against the C-terminus of Lyst, generated in collaboration with Dr. Stephen Brandt, did recognize a protein with the expected size for Lyst in normal mouse cells. However, this antibody recognized a protein with the same size in bg/bg cells which are not supposed to produce full-length Lyst because of a LINE1 element retrotransposition leading to a premature termination. Therefore, it appears that this antibody is not Lyst-specific and cannot be used for coimmunoprecipitation experiments. Generation of an anti-LYST antibody and confirmation of the LYST:LYST-IP interactions would be an important step in the future characterization of the Chediak-Higashi syndrome gene and its products.

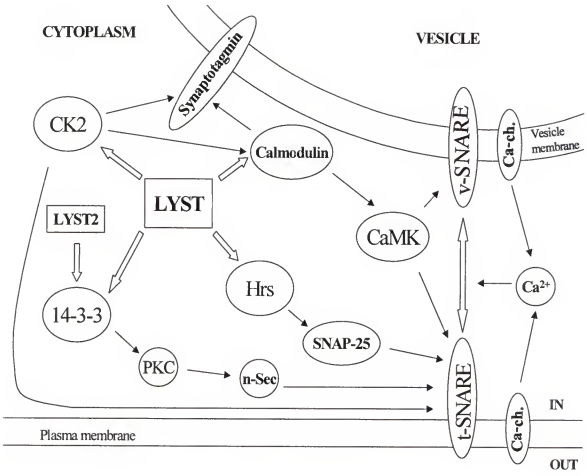

Figure 17. A simplified schematic representation of the interactions between LYST, LYST2, some of their interacting proteins and other relevant molecules in the context of neurotransmitter vesicle docking/fusion with the presynaptic plasma membrane of a neuron. All proteins form a common network of interacting molecules involved in the regulation of vesicle exocytosis.

In summary, it was demonstrated that each of the previously reported *bg* gene sequences are derived from a single gene with alternatively spliced mRNAs. Alternative splicing results in *bg* gene isoforms (*Lyst*-I and *Lyst*-II) that contain different 3' regions. Similarly to mouse, the human mRNA isoforms arise from incomplete splicing and retention of a transcribed intron that encodes the C-terminus of the predicted LYST protein. Additional splicing complexity of smaller isoforms exists - *Lyst*-III isoform lacks exons α and β, while in *Lyst*-IV, exons α, β and γ are absent. Novel mutations were identified within the coding domain of *LYST* in several CHS patients and *bg* alleles. Interestingly, all *bg* and *LYST* mutations identified to date are predicted to result in the production of either truncated or absent proteins. Mutation and expression analyses suggested that defects in the full-length mRNA alone can elicit CHS and that expression of the smaller isoform alone cannot compensate for loss of the largest isoform. We have also identified a novel gene in human (*LYST2)* and mouse (*Lyst2*) that appears to be a relative of the CHS gene, based on sequence similarity, predicted protein structure and on similar transcript size. Comparison of the relative abundance of *LYST2* and *LYST* mRNAs suggests that *LYST2* is expressed abundantly only in brain and therefore it may represent a brain-specific member of the *CHS* gene family. Using a modified yeast two-hybrid system, a human cDNA library was screened with baits from the coding domains of LYST and LYST2. Several proteins, which play important roles in protein transport and signal transduction, such as 14-3-3, CK2, calmodulin and Hrs, were found to interact with LYST and/or LYST2. Many of the interacting proteins could be linked in a common pathway that regulates vesicular trafficking and degranulation.

The molecular characterization of the Chediak-Higashi syndrome gene is of medical significance for several reasons. It will lead to improved understanding of CHS pathology and may suggest treatments that are not currently being considered. Targets for drugs that will limit degranulation and thereby ameliorate organ damage in inflammatory diseases may be developed as a result of characterization of molecules that regulate lysosomal secretion from granular cells. The observation that *bg* mice are protected from immune complex nephritis in the setting of systemic autoimmune disease (Clark *et al.*, 1982) suggests that selective blockade of Lyst may be a novel mode of protection from immunopathology in inflammatory diseases such as lupus nephritis. Finally, Lyst characterization may contribute to the molecular dissection of the associated cancer predisposition and clarify the role of natural immunity in surveillance against tumor development and metastasis.

APPENDIX
SEQUENCES OF MOLECULES INTERACTING WITH LYST AND LYST2

For all previously known proteins, GenBank accession numbers are shown in parentheses. The arrows indicate the beginning of the interacting fragments.

SEQUENCE 1. 14-3-3 PROTEIN (X56468)

```
  1  GTGGTGGGACTCGCGTCGCGGCCGCGGAGACGTGAAGCTCTCGAG
 46  GCTCCTCCCGCTGCGGGTCGGCGCTCGCCCTCGCTCTCCTCGCCC

 91  TCCGCCCCGGCCCCGGCCCCGGCCCCGCGCCCGCCATGGAGAAGA
                                      MetGluLysT

136  CTGAGCTGATCCAGAAGGCCAAGCTGGCCGAGCAGGCCGAGCGCT
     hrGluLeuIleGlnLysAlaLysLeuAlaGluGlnAlaGluArgT

181  ACGACGACATGGCCACCTGCATGAAGGCAGTGACCGAGCAGGGCG
     yrAspAspMetAlaThrCysMetLysAlaValThrGluGlnGlyA
                                                 ▼
226  CCGAGCTGTCCAACGAGGAGCGCAACCTGCTCTCCGTGGCCTACA
     laGluLeuSerAsnGluGluArgAsnLeuLeuSerValAlaTyrL
                  ▼                  ▼
271  AGAACGTGGTCGGGGGCCGCAGGTCCGCCTGGAGGGTCATCTCTA
     **ys**AsnValValGlyGly**Arg**ArgSerAlaTrpArg**Val**IleSerS
                  ▼                  ▼
316  GCATCGAGCAGAAGACCGACACCTCCGACAAGAAGTTGCAGCTGA
     erIleGluGlnLys**Thr**AspThrSerAspLysLeu**Gln**LeuI

361  TTAAGGACTATCGGGAGAAAGTGGAGTCCGAGCTGAGATCCATCT
     leLysAspTyrArgGluLysValGluSerGluLeuArgSerIleC

406  GCACCACGGTGCTGGAATTGTTGGATAAATATTTAATAGCCAATG
     ysThrThrValLeuGluLeuLeuAspLysTyrLeuIleAlaAsnA

451  CAACTAATCCAGAGAGTAAGGTCTTCTATCTGAAAATGAAGGGTG
     laThrAsnProGluSerLysValPheTyrLeuLysMetLysGlyA
```

```
496  ATTACTTCCGGTACCTTGCTGAAGTTGCGTGTGGTGATGATCGAA
     spTyrPheArgTyrLeuAlaGluValAlaCysGlyAspAspArgL

541  AACAAACGATAGATAATTCCCAAGGAGCTTACCAAGAGGCATTTG
     ysGlnThrIleAspAsnSerGlnGlyAlaTyrGlnGluAlaPheA

586  ATATAAGCAAGAAAGAGATGCAACCCACACACCCAATCCGCCTGG
     spIleSerLysLysGluMetGlnProThrHisProIleArgLeuG

631  GGCTTGCTCTTAACTTTTCTGTATTTTACTATGAGATTCTTAATA
     lyLeuAlaLeuAsnPheSerValPheTyrTyrGluIleLeuAsnA

676  ACCCAGAGCTTGCCTGCACGCTGGCTAAAACGGCTTTTGATGAGG
     snProGluLeuAlaCysThrLeuAlaLysThrAlaPheAspGluA

721  CCATTGCTGAACTTGATACACTGAATGAAGACTCATACAAAGACA
     laIleAlaGluLeuAspThrLeuAsnGluAspSerTyrLysAspS

766  GCACCCTCATCATGCAGTTGCTTAGAGACAACCTAACACTTTGGA
     erThrLeuIleMetGlnLeuLeuArgAspAsnLeuThrLeuTrpT

811  CATCAGACAGTGCAGGAGAAGAATGTGATGCGGCAGAAGGGGCTG
     hrSerAspSerAlaGlyGluGluCysAspAlaAlaGluGlyAlaG

856  AAAACTAAATCCATACAGGGTGTCATCCTTCTTTCCTTCAAGAAA
     luAsn

901  CCTTTTTACACATCTCCATTCCTTATTCCACTTGGATTTCCTATA
946  GCAAAGAAACCCATTCATGTGTATGGAATCAACTGTTTATAGTCT
991  TTTCACACTGCAGCTTTGGGAAAACTTCATTCCTTGATTTGTGTT
1036 TGTCTTGGCCTTCCTGGTGTGCAGTACTGCTGTAGAAAAGTATTA
1081 ATAGCTTCATTTCATATAAACATAAGTAACTCCCAAACACTTATG
1126 TAGAGGACTAAAAATGTATCTGGTATTTAAGTAATCTGAACCAGT
1171 TCTGCAAGTGACTGTGTTTTGTATTACTGTGAAAATAAGAAAATG
1216 TAGTTAATTACAATTTAAAGAGTATTCCACATAACTTCTTAATTT
1261 CTACATTCCCTCCCTTACTCTTCGGGGGTTTCCTTTCAGTAAGCA
1306 ACTTTTCCATGCTCTTAATGTATTCCTTTTTAGTAGGAATCCGGA
1351 AGTATTAGATTGAATGGAAAAGCACTTGCCATCTCTGTCTAGGGG
1396 TCACAAATTGAAATGGCTCCTGTATCACATACCGGAGGTCTTGTG
1441 TATCTGTGGCCAACAGGGAGTTTCCTTATTCACTCTTTATTTGCT
1486 GCTGTTTAAGTTGCCAACCTCCCCTCCCAATAAAAATTCACTTAC
1531 ACCTCCTGCCTTTGTAGTTCTGGTATTCACTTTACTATGTGATAG
1576 AAGTAGCATGTTGCTGCCAGAATACAAGCATTGCTTTTGGCAAAT
1621 TAAAGTGCATGTCATTTCTTAATACACTAGAAAGGGGAAATAAAT
1666 TAAAGTACACAAGTCCAAGTCTAAAACTTTAGTACTTTTCCCATG
1711 CAGATTTGTGCACATGTGAGAGGGTGTCCAGTTTGTCTAGTGATT
1756 GTTATTTAGAGAGTTGGACCACTATTGTGTGTTGCTAATCATTGA
1801 CTGTAGTCCCAAAAAAGCCTTGTGAAAATGTTATGCCCTATGTAA
1846 CAGCAGAGTAACATAAA
```

SEQUENCE 2. B14-3-3 PROTEIN (X57346)

```
   1  TACCGCCACCGCCGCCGCCGATTCCGGAGCCGGGGTAGTCGCCGC
  46  CGCCGCCGCCGCCGCTGCAGCCACTGCAGGCACCGCTGCCGCCGC
  91  CTGAGTAGTGTACCGCCACCGCCGCCGCCGATTCCGGAGCCGGGG
 136  TAGTCGCCGCCGCCGCCGCCGCCGCTGCAGCCACTGCAGGCACCG
 181  CTGCCGCCGCCTGAGTAGTGGGCTTAGGAAGGAAGAGGTCATCTC
 226  GCTCGGAGCTTCGCTCGGAAGGGTCTTTGTTCCCTGCAGCCCTCC
 271  CACGGCAGAGTCTCCAGAGATTTGGGCCGCTACAAAAAGTGCATT
 316  TTGCCCATTCGGCTGTGGATAGAGAAGCAGGAAGAGCACTGGACT

 361  TGGAGTCAGGGAATGACAATGGATAAAAGTGAGCTGGTACAGAAA
                    MetThrMetAspLysSerGluLeuValGlnLys

 406  GCCAAACTCGCTGAGCAGGCTGAGCGCTATGATGATATGGCTGCA
      AlaLysLeuAlaGluGlnAlaGluArgTyrAspAspMetAlaAla
                          ▼
 451  GCCATGAAGGCAGTCACAGAACAGGGGCATGAACTCTCCAACGAA
      AlaMetLysAlaVal**Thr**GluGlnGlyHisGluLeuSerAsnGlu

 496  GAGAGAAATCTGCTCTCTGTTGCCTACAAGAATGTGGTAGGCGCC
      GluArgAsnLeuLeuSerValAlaTyrLysAsnValValGlyAla

 541  CGCCGCTCTTCCTGGCGTGTCATCTCCAGCATTGAGCAGAAAACA
      ArgArgSerSerTrpArgValIleSerSerIleGluGlnLysThr

 586  GAGAGGAATGAGAAGAAGCAGCAGATGGGCAAAGAGTACCGTGAG
      GluArgAsnGluLysLysGlnGlnMetGlyLysGluTyrArgGlu

 631  AAGATAGAGGCAGAACTGCAGGACATCTGCAATGATGTTCTGGAG
      LysIleGluAlaGluLeuGlnAspIleCysAsnAspValLeuGlu

 676  CTGTTGGACAAATATCTTATTCCCAATGCTACACAACCAGAAAGT
      LeuLeuAspLysTyrLeuIleProAsnAlaThrGlnProGluSer

 721  AAGGTGTTCTACTTGAAAATGAAAGGAGATTATTTTAGGTATCTT
      LysValPheTyrLeuLysMetLysGlyAspTyrPheArgTyrLeu

 766  TCTGAAGTGGCATCTGGAGACAACAAACAAACCACTGTGTCGAAC
      SerGluValAlaSerGlyAspAsnLysGlnThrThrValSerAsn

 811  TCCCAGCAGGCTTACCAGGAAGCATTTGAAATTAGTAAGAAAGAA
      SerGlnGlnAlaTyrGlnGluAlaPheGluIleSerLysLysGlu

 856  ATGCAGCCTACACACCCAATTCGTCTTGGTCTGGCACTAAATTTC
      MetGlnProThrHisProIleArgLeuGlyLeuAlaLeuAsnPhe

 901  TCAGTCTTTTACTATGAGATTCTAAACTCTCCTGAAAAGGCCTGT
      SerValPheTyrTyrGluIleLeuAsnSerProGluLysAlaCys

 946  AGCCTGGCAAAAACGGCATTTGATGAAGCAATTGCTGAATTGGAT
      SerLeuAlaLysThrAlaPheAspGluAlaIleAlaGluLeuAsp
```

```
 991  ACGCTGAATGAAGAGTCTTATAAAGACAGCACTCTGATCATGCAG
      ThrLeuAsnGluGluSerTyrLysAspSerThrLeuIleMetGln

1036  TTACTTAGGGACAATCTCACTCTGTGGACATCGGAAAACCAGGGA
      LeuLeuArgAspAsnLeuThrLeuTrpThrSerGluAsnGlnGly

1081  GACGAAGGAGACGCTGGGGAGGGAGAGAACTAATGTTTCTCGTGC
      AspGluGlyAspAlaGlyGluGlyGluAsn

1126  TTTGTGATCTGTCCAGTGTCACTCTGTACCCTCAACATATATCCC
1171  TTGTGCGATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

SEQUENCE 3. CASEIN KINASE II β-SUBUNIT (M30448)

```
                1
    TCCGGGCCTGGCCTGTACGGTCGCCGCCGTTCCCTGGAAGTAGCA

46  ACTTCCCTACCCTACCCCAGTCCTGGTCCCCGTCCATCCGCTGAC

91  GTGAAGATGAGCAGCTCAGAGGAGGTGTCCTGGATTTCCTGGTTC
          MetSerSerSerGluGluValSerTrpIleSerTrpPhe

136 TGTGGGCTCCGTGGCAATGAATTCTTCTGTGAAGTGGATGAAGAC
    CysGlyLeuArgGlyAsnGluPhePheCysGluValAspGluAsp

181 TACATCCAGGACAAATTTAATCTTACTGGACTCAATGAGCAGGTC
    TyrIleGlnAspLysPheAsnLeuThrGlyLeuAsnGluGlnVal

226 CCTCACTACCGACAAGCTCTAGACATGATCTTGGACCTGGAGCCT
    ProHisTyrArgGlnAlaLeuAspMetIleLeuAspLeuGluPro

271 GATGAAGAACTGGAAGACAACCCCAACCAGAGTGACCTGATTGAG
    AspGluGluLeuGluAspAsnProAsnGlnSerAspLeuIleGlu

316 CAGGCAGCCGAGATGCTTTATGGATTGATCCACGCCCGCTACATC
    GlnAlaAlaGluMetLeuTyrGlyLeuIleHisAlaArgTyrIle

361 CTTACCAACCGTGGCATCGCCCAGATGTTGGAAAAGTACCAGCAA
    LeuThrAsnArgGlyIleAlaGlnMetLeuGluLysTyrGlnGln

406 GGAGACTTTGGTTACTGTCCTCGTGTGTACTGTGAGAACCAGCCA
    GlyAspPheGlyTyrCysProArgValTyrCysGluAsnGlnPro

451 ATGCTTCCCATTGGCCTTTCAGACATCCCAGGTGAAGCCATGGTG
    MetLeuProIleGlyLeuSerAspIleProGlyGluAlaMetVal

496 AAGCTCTACTGCCCCAAGTGCATGGATGTGTACACACCCAAGTCA
    LysLeuTyrCysProLysCysMetAspValTyrThrProLysSer

541 TCAAGACACCATCACACGGATGGCGCCTACTTCGGCACTGGTTTC
    SerArgHisHisHisThrAspGlyAlaTyrPheGlyThrGlyPhe

586 CCTCACATGCTCTTCATGGTGCATCCCGAGTACCGGCCCAAGAGA
    ProHisMetLeuPheMetValHisProGluTyrArgProLysArg

631 CCTGCCAACCAGTTTGTGCCCAGGCTCTACGGTTTCAAGATCCAT
    ProAlaAsnGlnPheValProArgLeuTyrGlyPheLysIleHis

676 GCGATGGCCTACCAGCTGCAGCTCCAAGCCGCCAGCAACTTCAAG
    AlaMetAlaTyrGlnLeuGlnLeuGlnAlaAlaSerAsnPheLys

721 AGCCCAGTCAAGACGATTCGCTGATTCCCTCCCCCACCTGTCCTG
    SerProValLysThrIleArg
```

766 CAGTCTTTGACTTTTCCTTTCTTTTTTGCCACCCTTTCAGGAACC
811 CTGTATGGTTTTTAGTTTAAATTAAAGGAGTCGTTATTGTGGTGG
856 GAATATGAAATAAAGTAGAAGAAAAGGCCATGAAAAAAAAAAAAA
901 AAAAAAAAAAAAAGGTTTGTGTTCTCGACGCCTTGGTAGTTGGCA
946 TAGGCTAAAGAAAAGGGATCTCAGCCCCGAGGAAGGGTCACCCTC
991 CTAGAGATAGCTACTACCCCGTCTCAGGAGACCCTGGTATTTCTA
1036 GAGCACGCTTTGCTTTCACCAAACCCAAGGAGGTGACAGGAGGAG
1081 CCCCCGCACAGGACCTAAGAATGCTGTGACCAGAAGATGGGATCG
1126 CGGAACAGCAGCAGTGCAGGATCCGGTCCGGAGACCCCTCCGAGG
1171 GCTTGCCCCGAAGAGGGGCTGGCCTGCGTCGGAGTGAGGAAGAGG
1216 AAGAAGAGGATGAAGATGTGGATCTGGCCCAGGTACTGGCCTATC
1261 TCCTCCGCAGAGGCCAAGTGAGGTTGGTGCAGGGAGGAGGTGCAG
1306 CAAATTTACAATTCATTCAGGCCCTCTTGGACTCAGAGGAAGAGA
1351 ATGACAGAGCTTGGGATGGTCGTCTTGGGGATCGATACAACCCAC
1396 CTGTGGATGCTACCCCTGACACCCGGGAGCTGGAATTCCTGGAGT
1441 TGTGAACGCCGCGGACTCCGGAGCCGCACAAACCAGGGCTCGCCA
1486 TGAAGCCAGGATTCAGTCCCCCGTGGGGGTGGCTTTGGCGGCCGA
1531 GGGGGCTTTGGTGACCGTGGTGGTCGTGGAGGCCGAGGGGGCTTT
1576 GGCGGGGGCCGAGGTCGAGGCGGAGGCTTTAGAGGTCGTGGACGA
1621 GGAGGAGGTGGAGGCGGCGGCGGCGGTGGAGGAGGAGGAAGAGGT
1666 GGAGGCTTCCATTCTGGTGGCAACCGGGGTCGTGGTCGGGGAGGA
1711 AAAAGAGGAAACCAGTCGGGGAAGAATGTGATGGTGGAGCCGCAT
1756 CGGACTGAGGGTGTCTTCATTTGTCGAGGAAAGGAGATGCACTGG
1801 TCACCAAGAACCTGGTCCTGGGAATCAAGTTATGGAGAGAAGAGA
1846 GTCTCGATTCGGAAGGAGATGACAAAATTGAGTACCGAGCCTGGA
1891 ACCCCTTCCGCTCCAAGCTAGCAGCAGCAATCCTGGGTGGTGTGG
1936 ACCAGATCCACATCAAACCGGGGGCTAAGGTTCTCTACCTCGGGG
1981 CTGCCTCGGGCACCACGGTCTCCCATGTCTCTGACATCGTTGGTC
2026 CGGATGGTCTAGTCTATGCAGTCGAGTTTCCCCACCGCTCTGGCC
2071 GTGACCTCATTAACTTGGCCAAGAAGAGGACCAACATCATTCCTG
2116 TGATCGAGGATGCTCGACACCCACACAAATACCGCATGCTCATCG
2161 CAATGGTGGATGTGATCTTTGCTGATGTGGCCCAGCCAGACCAGA
2206 CCCGGAATGTGGCCCTGAAAGCCCACACCTTCCTGCGCTAATGGA
2251 GGACACTTTGTGATTTCCATTAAGGCCAACTGCATTGACTCCACA
2296 GCCTAGCGGTCAGGCCGTGTTTGCCTCCGAAGTGAAAAAGATGCA
2341 ACAGGAGAACATGAAGCCGCAGGAGCAGTTGACCCTTGAGCCATA
2386 TGAAAGAGACCATGCCGTGGTCGTGGGAGTGTACAGGCCACCCCC
2431 CAAGGTGAAGAACTGAAGTTCAGCGCTGTCAGGATTGCGAGAGAT
2476 GTGTGTTGATACTGTTGCCCGTGTGTTTTTCTAATTAAAAGACTC
2521 CTCCGTC

SEQUENCE 4. HEPATOCYTE GROWTH FACTOR-REGULATED TYROSINE KINASE SUBSTRATE (HRS, D84064)

```
1   GGGCGCGCCAGCTCGTAGCAGGGGAGCGCCCGCGGCGTCGGGTTT

46  GGGCTGGAGGTCGCCATGGGGCGAGGCAGCGGCACCTTCGAGCGT
                  MetGlyArgGlySerGlyThrPheGluArg

91  CTCCTAGACAAGGCGACCAGCCAGCTCCTGTTGGAGACAGATTGG
    LeuLeuAspLysAlaThrSerGlnLeuLeuLeuGluThrAspTrp

136 GAGTCCATTTTGCAGATCTGCGACCTGATCCGCCAAGGGGACACA
    GluSerIleLeuGlnIleCysAspLeuIleArgGlnGlyAspThr

181 CAAGCAAAATATGCTGTGAATTCCATCAAGAAGAAAGTCAACGAC
    GlnAlaLysTyrAlaValAsnSerIleLysLysLysValAsnAsp

226 AAGAACCCACACGTCGCCTTGTATGCCCTGGAGGTCATGGAATCT
    LysAsnProHisValAlaLeuTyrAlaLeuGluValMetGluSer
                                        ▼
271 GTGGTAAAGAACTGTGGCCAGACAGTTCATGATGAGGTGGCCAAC
    ValValLysAsnCysGlyGlnThrValHisAspGluVal**Ala**Asn

316 AAGCAGACCATGGAGGAGCTGAAGGACCTGCTGAAGAGACAAGTG
    LysGlnThrMetGluGluLeuLysAspLeuLeuLysArgGlnVal

361 GAGGTAAACGTCCGTAACAAGATCCTGTACCTGATCCAGGCCTGG
    GluValAsnValArgAsnLysIleLeuTyrLeuIleGlnAlaTrp

406 GCGCATGCCTTCCGGAACGAGCCCAAGTACAAGGTGGTCCAGGAC
    AlaHisAlaPheArgAsnGluProLysTyrLysValValGlnAsp

451 ACCTACCAGATCATGAAGGTGGAGGGGCACGTCTTTCCAGAATTC
    ThrTyrGlnIleMetLysValGluGlyHisValPheProGluPhe

496 AAAGAGAGCGATGCCATGTTTGCTGCCGAGAGAGCCCCAGACTGG
    LysGluSerAspAlaMetPheAlaAlaGluArgAlaProAspTrp

541 GTGGACGCTGAGGAATGCCACCGCTGCAGGGTGCAGTTCGGGGTG
    ValAspAlaGluGluCysHisArgCysArgValGlnPheGlyVal

586 ATGACCCGTAAGCACCACTGCCGGGCGTGTGGGCAGATATTCTGT
    MetThrArgLysHisHisCysArgAlaCysGlyGlnIlePheCys

631 GGAAAGTGTTCTTCCAAGTACTCCACCATCCCCAAGTTTGGCATC
    GlyLysCysSerSerLysTyrSerThrIleProLysPheGlyIle

676 GAGAAGGAGGTGCGCGTGTGTGAGCCCTGCTACGAGCAGCTGAAC
    GluLysGluValArgValCysGluProCysTyrGluGlnLeuAsn

721 AGGAAAGCGGAGGGAAAGGCCACTTCCACCACTGAGCTGCCCCCC
    ArgLysAlaGluGlyLysAlaThrSerThrThrGluLeuProPro
```

```
766  GAGTACCTGACCAGCCCCCTGTCTCAGCAGTCCCAGCTGCCCCCC
     GluTyrLeuThrSerProLeuSerGlnGlnSerGlnLeuProPro

811  AAGAGGGACGAGACGGCCCTGCAGGAGGAGGAGGAGCTGCAGCTG
     LysArgAspGluThrAlaLeuGlnGluGluGluGluLeuGlnLeu

856  GCCCTGGCGCTGTCACAGTCAGAGGCGGAGGAGAAGGAGAGGCTG
     AlaLeuAlaLeuSerGlnSerGluAlaGluGluLysGluArgLeu

901  AGACAGAAGTCCACGTACACTTCGTACCCCAAGGCGGAGCCCATG
     ArgGlnLysSerThrTyrThrSerTyrProLysAlaGluProMet

946  CCCTCGGCCTCCTCAGCGCCCCCCGCCAGCAGCCTGTACTCTTCA
     ProSerAlaSerSerAlaProProAlaSerSerLeuTyrSerSer

991  CCTGTGAACTCGTCGGCGCCTCTGGCTGAGGACATCGACCCTGAG
     ProValAsnSerSerAlaProLeuAlaGluAspIleAspProGlu

1036 CTCGCACGGTATCTCAACCGGAACTACTGGGAGAAGAAGCAGGAG
     LeuAlaArgTyrLeuAsnArgAsnTyrTrpGluLysLysGlnGlu

1081 GAGGCTCGCAAGAGCCCCACGCCATCTGCGCCCGTGCCCCTGACG
     GluAlaArgLysSerProThrProSerAlaProValProLeuThr

1126 GAGCCGGCTGCACAGCCTGGGGAAGGGCACGCAGCCCCCACCAAC
     GluProAlaAlaGlnProGlyGluGlyHisAlaAlaProThrAsn

1171 GTGGTGGAGAACCCCCTCCCGGAGACAGACTCTCAGCCCATTCCT
     ValValGluAsnProLeuProGluThrAspSerGlnProIlePro

1216 CCCTCTGGTGGCCCCTTTAGTGAGCCACAGTTCCACAATGGCGAG
     ProSerGlyGlyProPheSerGluProGlnPheHisAsnGlyGlu

1261 TCTGAGGAGAGCCACGAGCAGTTCCTGAAGGCGCTGCAGAACGCC
     SerGluGluSerHisGluGlnPheLeuLysAlaLeuGlnAsnAla

1306 GTCACCACCTTCGTGAACCGCATGAAGAGTAACCACATGCGGGGC
     ValThrThrPheValAsnArgMetLysSerAsnHisMetArgGly

1351 CGCAGCATCACCAATGACTCGGCCGTGCTCTCACTCTTCCAGTCC
     ArgSerIleThrAsnAspSerAlaValLeuSerLeuPheGlnSer

1396 ATCAACGGCATGCACCCGCAGCTGCTGGAGCTGCTCAACCAGCTG
     IleAsnGlyMetHisProGlnLeuLeuGluLeuLeuAsnGlnLeu

1441 GACGAGCGCAGGCTGTACTATGAGGGGCTGCAGGACAAGCTGGCA
     AspGluArgArgLeuTyrTyrGluGlyLeuGlnAspLysLeuAla

1486 CAGATCCGCGATGCCCGGGGGGCGCTGAGTGCCCTGCGCGAAGAG
     GlnIleArgAspAlaArgGlyAlaLeuSerAlaLeuArgGluGlu

1531 CACCGGGAGAAGCTTCGCCGGGCAGCCGAGGAGGCAGAGCGCCAG
     HisArgGluLysLeuArgArgAlaAlaGluGluAlaGluArgGln
```

```
1576  CGCCAGATCCAGCTGGCCCAGAAGCTGGAGATAATGCGGCAGAAG
      ArgGlnIleGlnLeuAlaGlnLysLeuGluIleMetArgGlnLys

1621  AAGCAGGAGTACCTGGAGGTGCAGAGGCAGCTGGCCATCCAGCGC
      LysGlnGluTyrLeuGluValGlnArgGlnLeuAlaIleGlnArg

1666  CTGCAGGAGCAGGAGAAGGAGCGGCAGATGCGGCTGGAGCAGCAG
      LeuGlnGluGlnGluLysGluArgGlnMetArgLeuGluGlnGln

1711  AAGCAGACGGTCCAGATGCGCGGCGACGATGCCCGCCTTCCCCCTG
      LysGlnThrValGlnMetArgAlaGlnMetProAlaProLeu

1756  CCCTACGCCCAGCTCCAGGCCATGCCCGCAGCCGGAGGTGTGCTC
      ProTyrAlaGlnLeuGlnAlaMetProAlaAlaGlyGlyValLeu

1801  TACCAGCCCTCGGGACCAGCCAGCTTCCCCAGCACCTTCAGCCCT
      TyrGlnProSerGlyProAlaSerProSerThrPheSerPro

1846  GCCGGCTCGGTGGAGGGCTCCCCAATGCACGGCGTGTACATGAGC
      AlaGlySerValGluGlySerProMetHisGlyValTyrMetSer

1891  CAGCCGGCCCCTGCCGCTGGCCCCTACCCCAGCATGCCCAGCACT
      GlnProAlaProAlaAlaGlyProTyrProSerMetProSerThr

1936  GCGGCTGATCCCAGCATGGTGAGTGCCTACATGTACCCAGCAGGG
      AlaAlaAspProSerMetValSerAlaTyrMetTyrProAlaGly

1981  GCCACTGGGGCGCAGGCGGCCCCCCAGGCCCAGGCCGGACCCACC
      AlaThrGlyAlaGlnAlaAlaProGlnAlaGlnAlaGlyProThr

2026  GCCAGCCCCGCTTACTCATCCTACCAGCCTACTCCCACAGCGGGC
      AlaSerProAlaTyrSerSerTyrGlnProThrProThrAlaGly

2071  TACCAGAACGTGGCCTCCCAGGCCCCACAGAGCCTCCCGGCCATC
      TyrGlnAsnValAlaSerGlnAlaProGlnSerLeuProAlaIle

2116  TCTCAGCCTCCGCAGTCCAGCACCATGGGCTACATGGGGAGCCAG
      SerGlnProProGlnSerSerThrMetGlyTyrMetGlySerGln

2161  TCAGTCTCCATGGGCTACCAGCCTTACAACATGCAGAATCTCATG
      SerValSerMetGlyTyrGlnProTyrAsnMetGlnAsnLeuMet

2206  ACCACCCTCCCAAGCCAGGATGCGTCTCTGCCACCCCAGCAGCCC
      ThrThrLeuProSerGlnAspAlaSerLeuProProGlnGlnPro

2251  TACATCGCGGGGCAGCAGCCCATGTACCAGCAGATGGCACCCTCT
      TyrIleAlaGlyGlnGlnProMetTyrGlnGlnMetAlaProSer

2296  GGCGGTCCCCCCCAGCAGCAGCCCCCCGTGGCCCAGCAACCGCAG
      GlyGlyProProGlnGlnGlnProProValAlaGlnGlnProGln

2341  GCACAGGGGCCGCCGGCACAGGGCAGCGAGGCCCAGCTCATTTCA
      AlaGlnGlyProProAlaGlnGlySerGluAlaGlnLeuIleSer
```

```
2386  TTCGACTGACCCAGGCCATGCTCACGTCCGGAGTAACACTACATA
      PheAsp

2431  CAGTTCACCTGAAACGCCTCGTCTCTAACTGCCGTCGTCCTGCCT
2476  CCCTGTCCTCTACTGCCGGTAGTGTCCCTTCTCTGCGAGTGAGGG
2521  GGGGCCTTCACCCCAAGCCCACCTCCCTTGTCCTCAGCCTACTGC
2566  AGTCCCTGAGTTAGTCTCTGCTTTCTTTCCCCAGGGCTGGGCCAT
2611  GGGGAGGGAAGGACTTTCTCCCAGGGGAAGCCCCCAGCCCTGTGG
2656  GTCATGGTCTGTGAGAGGTGGCAGGAATGGGGACCCTCACCCCCC
2701  AAGCAGCCTGTGCCCTCTGGCCGCACTGTGAGCTGGCTGTGGTGT
2746  CTGGGTGTGGCCTGGGGCTCCCTCTGCAGGGGCCTCTCTCGGCAG
2791  CCACAGCCAAGGGTGGAGGCTTCAGGTCTCCAGCTTCTCTGCTTC
2836  TCAGCTGCCATCTCCAGTGCCCCAGAATGGTACAGCGATAATAAA
2881  ATGTATTTCAGAAAGG
```

SEQUENCE 5. CALMODULIN (D45887)

```
   1 AGTCCGAGTGGAGAGAGCGAGCTGAGTGGTTGTGTGGTCGCGTCT

  46 CGGAAACCGGTAGCGCTTGCAGCATGGCTGACCAACTGACTGAAG
                              MetAlaAspGlnLeuThrGluG

  91 AGCAGATTGCAGAATTCAAAGAAGCTTTTTCACTATTTGACAAAG
     luGlnIleAlaGluPheLysGluAlaPheSerLeuPheAspLysA

 136 ATGGTGATGGAACTATAACAACAAAGGAATTGGGAACTGTAATGA
     spGlyAspGlyThrIleThrThrLysGluLeuGlyThrValMetA

 181 GATCTCTTGGGCAGAATCCCACAGAAGCAGAGTTACAGGACATGA
     rgSerLeuGlyGlnAsnProThrGluAlaGluLeuGlnAspMetI

 226 TTAATGAAGTAGATGCTGATGGTAATGGCACAATTGACTTCCCTG
     leAsnGluValAspAlaAspGlyAsnGlyThrIleAspPheProG

 271 AATTTCTGACAATGATGGCAAGAAAAATGAAAGACACAGACAGTG
     luPheLeuThrMetMetAlaArgLysMetLysAspThrAspSerG

 316 AAGAAGAAATTAGAGAAGCATTCCGTGTGTTTGATAAGGATGGCA
     luGluGluIleArgGluAlaPheArgValPheAspLysAspGlyA

 361 ATGGCTATATTAGTGCTGCAGAACTTCGCCATGTGATGACAAACC
     snGlyTyrIleSerAlaAlaGluLeuArgHisValMetThrAsnL

 406 TTGGAGAGAAGTTAACAGATGAAGAAGTTGATGAAATGATCAGGG
     euGlyGluLysLeuThrAspGluGluValAspGluMetIleArgG
                                 ▼
 451 AAGCAGATATTGATGGTGATGGTCAAGTAAACTATGAAGAGTTTG
     luAlaAspIleAspGlyAspGlyGlnValAsnTyrGluGluPheV

 496 TACAAATGATGACAGCAAAGTGAAGACCTTGTACAGAATGTGTTA
     alGlnMetMetThrAlaLys
 541 AATTTCTTGTACAAAATTGTTTATTTGCCTTTTCTTTGTTTGTAA
 586 CTTATCTGTAAAAGGTTTCTCCCTACTGTCAAAAAAATATGCATG
 631 TATAGTAATTAGGACTTCATTCCTCCATGTTTTCTTCCCTTATCT
 676 TACTGTCATTGTCCTAAAACCTTATTTTAGAAAATTGATCAAGTA
 721 ACATGTTGCATGTGGCTTACTCTGGATATATCTAAGCCCTTCTGC
 766 ACATCTAAACTTAGATGGAGTTGGTCAAATGAGGGAACATCTGGG
 811 TTATGCCTTTTTTAAAGTAGTTTTCTTTAGGAACTGTCAGCATGT
 856 TGTTGTTGAAGTGTGGAGTTGTAACTCTGCGTGGACTATGGACAG
 901 TCAACAATATGTACTTAAAAGTTGCACTATTGCAAAACGGGTGTA
 946 TTATCCAGGTACTCGTACACTATTTTTTTGTACTGCTGGTCCTGT
 991 ACCAGAAACATTTTCTTTTATTGTTACTTGCTTTTTAAACTTTGT
1036 TTAGCCACTTAAAATCTGCTTATGGCACAATTTGCCTCAAAATCC
1081 ATTCCAAGTTGTATATTTGTTTTCCAATAAAAAAATTACAATTTA
1126 CCC
```

SEQUENCE 6. ESTROGEN RECEPTOR-RELATED PROTEIN (L38487)

```
   1 TCCTACAAGCAGCCGGCGGCGCCGCCGAGTGAGGGGACGCGGCGC
     SerTyrLysGlnProAlaAlaProProSerGluGlyThrArgArg

  46 GGTGGGGCGGCGCGGCCCGAGGAGGCGGCGGAGGAGGGGCCGCCC
     GlyGlyAlaAlaArgProGluGluAlaAlaGluGluGlyProPro

  91 GCGGCCCCCGGCTCACTCCGGCACTCCGGGCCGCTCGGCCCCCAT
     AlaAlaProGlySerLeuArgHisSerGlyProLeuGlyProHis

 136 GCCTGCCCGACCGCGCTGCCGGAGCCCCAGGTGACCAGCGCCATG
     AlaCysProThrAlaLeuProGluProGlnValThrSerAlaMet

 181 TCCAGCCAGGTGGTGGGCATTGAGCCTCTCTACATCAAGGCAGAG
     SerSerGlnValValGlyIleGluProLeuTyrIleLysAlaGlu

 226 CCGGCCAGCCCTGACAGTCCAAAGGGTTCCTCGGAGACAGAGACC
     ProAlaSerProAspSerProLysGlySerSerGluThrGluThr

 271 GAGCCTCCTGTGGCCCTGGCCCCTGGTCCAGCTCCCACTCGCTGC
     GluProProValAlaLeuAlaProGlyProAlaProThrArgCys

 316 CTCCCAGGCCACAAGGAAGAGGAGGATGGGGAGGGGGCTGGGCCT
     LeuProGlyHisLysGluGluGluAspGlyGluGlyAlaGlyPro

 361 GGCGAGCAGGGCGGTGGGAAGCTGGTGCTCAGCTCCCTGCCCAAG
     GlyGluGlnGlyGlyGlyLysLeuValLeuSerSerLeuProLys

 406 CGCCTCTGCCTGGTCTGTGGGGACGTGGCCTCCGGCTACCACTAT
     ArgLeuCysLeuValCysGlyAspValAlaSerGlyTyrHisTyr

 451 GGTGTGGCATCCTGTGAGGCCTGCAAAGCCTTCTTCAAGAGGACC
     GlyValAlaSerCysGluAlaCysLysAlaPhePheLysArgThr

 496 ATCCAGGGGAGCATCGAGTACAGCTGTCCGGCCTCCAACGAGTGT
     IleGlnGlySerIleGluTyrSerCysProAlaSerAsnGluCys

 541 GAGATCACCAAGCGGAGACGCAAGGCCTGCCAGGCCTGCCGCTTC
     GluIleThrLysArgArgArgLysAlaCysGlnAlaCysArgPhe

 586 ACCAAGTGCCTGCGGGTGGGCATGCTCAAGGAGGGAGTGCGCCTG
     ThrLysCysLeuArgValGlyMetLeuLysGluGlyValArgLeu

 631 GACCGCGTCCGGGGTGGGCGGCAGAAGTACAAGCGGCGGCCGGAG
     AspArgValArgGlyGlyArgGlnLysTyrLysArgArgProGlu

 676 GTGGACCCACTGCCCTTCCCGGGCCCCTTCCCTGCTGGGCCCCTG
     ValAspProLeuProProGlyProPheProAlaGlyProLeu
                                    ▼
 721 GCAGTCGCTGGAGGCCCCCGGAAGACAGCCCCAGTGAATGCACTG
     AlaValAlaGlyGlyProArgLysThrAlaProVal**Asn**AlaLeu
```

```
 766 GTGTCTCATCTGCTGGTGGTTGAGCCTGAGAAGCTCTATGCCATG
     ValSerHisLeuLeuValValGluProGluLysLeuTyrAlaMet

 811 CCTGACCCCGCAGGCCCTGATGGGCACCTCCCAGCCGTGGCTACC
     ProAspProAlaGlyProAspGlyHisLeuProAlaValAlaThr

 856 CTCTGTGACCTCTTTGACCGAGAGATTGTGGTCACCATCAGCTGG
     LeuCysAspLeuPheAspArgGluIleValValThrIleSerTrp

 901 GCCAAGAGCATCCCAGGCTTCTCATCGCTGTCGCTGTCTGACCAG
     AlaLysSerIleProGlyPheSerSerLeuSerLeuSerAspGln

 946 ATGTCAGTACTGCAGAGCGTGTGGATGGAGGTGCTGGTGCTGGGT
     MetSerValLeuGlnSerValTrpMetGluValLeuValLeuGly

 991 GTGGCCCAGCGCTCACTGCCACTGCAGGATGAGCTGGCCTTCGCT
     ValAlaGlnArgSerLeuProLeuGlnAspGluLeuAlaPheAla

1036 GAGGACTTAGTCCTGGATGAAGAGGGGGCACGGGCAGCTGGCCTG
     GluAspLeuValLeuAspGluGluGlyAlaArgAlaAlaGlyLeu

1081 GGGGAACTGGGGGCTGCCCTGCTGCAACTAGTGCGGCGGCTGCAG
     GlyGluLeuGlyAlaAlaLeuLeuGlnLeuValArgArgLeuGln

1126 GCCCTGCGGCTGGAGCGAGAGGAGTATGTTCTACTAAAGGCCTTG
     AlaLeuArgLeuGluArgGluGluTyrValLeuLeuLysAlaLeu

1171 GCCCTTGCCAATTCAGACTCTGTGCACATCGAAGATGCCGAGGCT
     AlaLeuAlaAsnSerAspSerValHisIleGluAspAlaGluAla

1216 GTGGAGCAGCTGCGAGAAGCTCTGCACGAGGCCCTGCTGGAGTAT
     ValGluGlnLeuArgGluAlaLeuHisGluAlaLeuLeuGluTyr

1261 GAAGCCGGCCGGGCTGGCCCCGGAGGGGGTGCTGAGCGGCGGCGG
     GluAlaGlyArgAlaGlyProGlyGlyGlyAlaGluArgArgArg

1306 GCGGGCAGGCTGCTGCTCACGCTACCGCTCCTCCGCCAGACAGCG
     AlaGlyArgLeuLeuLeuThrLeuProLeuLeuArgGlnThrAla

1351 GGCAAAGTGCTGGCCCATTTCTATGGGGTGAAGCTGGAGGGCAAG
     GlyLysValLeuAlaHisPheTyrGlyValLysLeuGluGlyLys

1396 GTGCCCATGCACAAGCTGTTCTTGGAGATGCTCGAGGCCATGATG
     ValProMetHisLysLeuPheLeuGluMetLeuGluAlaMetMet

1441 GACTGAGGCAAGGGGTGGGACTGGTGGGGGTTCTGGCAGGACCTG
     Asp

1486 CCTAGCATGGGGTCAGCCCCAAGGGCTGGGGCGGAGCTGGGGTCT
1531 GGGCAGTGCCACAGCCTGCTGGCAGGGCCAGGGCAATGCCATCAG
1576 CCCCTGGGACAGGCCCCACGCCCTCTCCTCCCCCTCCTAGGGGG
1621 TGTCAGAAGCTGGGAACGTGTGTCCAGGCTCTGGGCACAGTGCTG
1666 CCCCTTGCAAGCCATAACGTGCCCCCAGAGTGTAGGGGGCCTTGC
```

```
1711 GGAAGCCATAGGGGGCTGCACGGGATGCGTGGGAGGCAGAAACCT
1756 ATCTCAGGGAGGGAAGGGGATGGAGGCCAGAGTCTCCCAGTGGGT
1801 GATGCTTTTGCTGCTGCTTAATCCTACCCCCTCTTCAAAGCAGAG
1846 TGGGACTTGGAGAGCAAAGGCCCATGCCCCCTTCGCTCCTCCTCT
1891 CATCATTTGCATTGGGCATTAGTGTCCCCCCTTGAAGCAATAACT
1936 CCAAGCAGACTCCAGCCCCTGGACCCCTGGGGTGGCCAGGGCTTC
1981 CCCATCAGCTCCCAACGAGCCTCCTCAGGGGGTAGGAGAGCACTG
2026 CCTCTATGCCCTGCAGAGCAATAACACTATATTTATTTTGGGTT
2071 TGGCCAGGGAGGCGCAGGGACATGGGGCAAGCCAGGGCCCAGAGC
2116 CCTTGGCTGTACAGAGACTCTATTTTAATGTATATTTGCTGCAAA
2161 GAGAAACCGCTTTTGGTTTTAAACCTTTAATGAGAAAAAATATA
2206 TAATACCGAGCTC
```

SEQUENCE 7. IMPORTIN β-SUBUNIT (L38951)

```
   1 CTCCCTCGCTCCCTCCCTGCGCGCCGCCTCTCACTCACAGCCTCC
  46 CTTCCTTCTTTCTCCCTCCGCCTCCCGAGCACCAGCGCGCTCTGA
  91 GCTGCCCCCAGGGTCCCCTCCCCCGCCGCCAGCAGCCCATTTGGA
 136 GGGAGGAAGTAAGGGAAGAGGAGAGGAAGGGGAGCCGGACCGACT
 181 ACCCAGACAGAGCCGGTGAATGGGTTTGTGGTGACCCCCGCCCCC
 226 CACCCCACCCTCCCTTCCCACCCGACCCCCCAACCCCCATCCCCAG
 271 TTCGAGCCGCCGCCCGAAAGGCCGGGCCGTCGTCTTAGGAGGAGT
 316 CGCCGCCGCCGCCACCTCCGCCATGGAGCTGATCACCATTCTCGA
                              MetGluLeuIleThrIleLeuGl

 361 GAAGACCGTGTCTCCCGATCGGCTGGAGCTGGAAGCGGCGCAGAA
     uLysThrValSerProAspArgLeuGluLeuGluAlaAlaGlnLy

 406 GTTCCTGGAGCGTGCGGCCGTGGAGAACCTGCCCACTTTCCTTGT
     sPheLeuGluArgAlaAlaValGluAsnLeuProThrPheLeuVa

 451 GGAACTGTCCAGAGTGCTGGCAAATCCAGGAAACAGTCAGGTTGC
     lGluLeuSerArgValLeuAlaAsnProGlyAsnSerGlnValAl

 496 CAGAGTTGCAGCTGGTCTACAAATCAAGAACTCTTTGACATCTAA
     aArgValAlaAlaGlyLeuGlnIleLysAsnSerLeuThrSerLy

 541 AGATCCAGATATCAAGGCACAATATCAGCAGAGGTGGCTTGCTAT
     sAspProAspIleLysAlaGlnTyrGlnGlnArgTrpLeuAlaIl

 586 TGATGCTAATGCTCGACGAGAAGTCAAGAACTATGTTTGCAGAC
     eAspAlaAsnAlaArgArgGluValLysAsnTyrValLeuGlnTh

 631 ATTGGGTACAGAAACTTACCGGCCTAGTTCTGCCTCACAGTGTGT
     rLeuGlyThrGluThrTyrArgProSerSerAlaSerGlnCysVa

 676 GGCTGGTATTGCTTGTGCAGAGATCCCAGTAAACCAGTGGCCAGA
     lAlaGlyIleAlaCysAlaGluIleProValAsnGlnTrpProGl

 721 ACTCATTCCTCAGCTGGTGGCCAATGTCACAAACCCCAACAGCAC
     uLeuIleProGlnLeuValAlaAsnValThrAsnProAsnSerTh

 766 AGAGCACATGAAGGAGTCGACATTGGAAGCCATCGGTTATATTTG
     rGluHisMetLysGluSerThrLeuGluAlaIleGlyTyrIleCy

 811 CCAAGATATAGACCCAGAGCAGCTACAAGATAAATCCAATGAGAT
     sGlnAspIleAspProGluGlnLeuGlnAspLysSerAsnGluIl

 856 TCTGACTGCCATAATCCAGGGGATGAGGAAAGAAGAGCCTAGTAA
     eLeuThrAlaIleIleGlnGlyMetArgLysGluGluProSerAs

 901 TAATGTGAAGCTAGCTGCTACAAATGCACTCCTGAACTCATTGGA
     nAsnValLysLeuAlaAlaThrAsnAlaLeuLeuAsnSerLeuGl

 946 GTTCACCAAAGCAAACTTTGATAAAGAGTCTGAAAGGCACTTTAT
     uPheThrLysAlaAsnPheAspLysGluSerGluArgHisPheIl
```

```
 991 TATGCAGGTGGTCTGTGAAGCCACACAGTGTCCAGATACGAGGGT
     eMetGlnValValCysGluAlaThrGlnCysProAspThrArgVa

1036 ACGAGTGGCTGCTTTACAGAATCTGGTGAAGATAATGTCCTTATA
     lArgValAlaAlaLeuGlnAsnLeuValLysIleMetSerLeuTy

1081 TTATCAGTACATGGAGACATATATGGGTCCTGCTCTTTTTGCAAT
     rTyrGlnTyrMetGluThrTyrMetGlyProAlaLeuPheAlaIl

1126 CACAATCGAAGCAATGAAAAGTGACATTGATGAGGTGGCTTTACA
     eThrIleGluAlaMetLysSerAspIleAspGluValAlaLeuGl

1171 AGGGATAGAATTCTGGTCCAATGTCTGTGATGAGGAAATGGATTT
     nGlyIleGluPheTrpSerAsnValCysAspGluGluMetAspLe

1216 GGCCATTGAAGCTTCAGAGGCAGCAGAACAAGGACGGCCCCCTGA
     uAlaIleGluAlaSerGluAlaAlaGluGlnGlyArgProProGl

1261 GCACACCAGCAAGTTTTATGCGAAGGGAGCACTACAGTATCTGGT
     uHisThrSerLysPheTyrAlaLysGlyAlaLeuGlnTyrLeuVa

1306 TCCAATCCTCACACAGACACTAACTAAACAGGACGAAAATGATGA
     lProIleLeuThrGlnThrLeuThrLysGlnAspGluAsnAspAs

1351 TGACGATGACTGGAACCCCTGCAAAGCAGCAGGGGTGTGCCTCAT
     pAspAspAspTrpAsnProCysLysAlaAlaGlyValCysLeuMe

1396 GCTTCTGGCCACCTGCTGTGAAGATGACATTGTCCCACATGTCCT
     tLeuLeuAlaThrCysCysGluAspAspIleValProHisValLe

1441 CCCCTTCATTAAAGAACACATCAAGAACCCAGATTGGCGGTACCG
     uProPheIleLysGluHisIleLysAsnProAspTrpArgTyrAr

1486 GGATGCAGCAGTGATGGCTTTTGGTTGTATCTTGGAAGGACCAGA
     gAspAlaAlaValMetAlaPheGlyCysIleLeuGluGlyProGl

1531 GCCCAGTCAGCTCAAACCACTAGTTATACAGGCTATGCCCACCCT
     uProSerGlnLeuLysProLeuValIleGlnAlaMetProThrLe

1576 AATAGAATTAATGAAAGACCCCAGTGTAGTTGTTCGAGATACAGC
     uIleGluLeuMetLysAspProSerValValValArgAspThrAl

1621 TGCATGGACTGTAGGCAGAATTTGTGAGCTGCTTCCTGAAGCTGC
     aAlaTrpThrValGlyArgIleCysGluLeuLeuProGluAlaAl

1666 CATCAATGATGTCTACTTGGCTCCCCTGCTACAGTGTCTGATTGA
     aIleAsnAspValTyrLeuAlaProLeuLeuGlnCysLeuIleGl

1711 GGGTCTCAGTGCTGAACCCAGAGTGGCTTCAAATGTGTGCTGGGC
     uGlyLeuSerAlaGluProArgValAlaSerAsnValCysTrpAl

1756 TTTCTCCAGTCTGGCTGAAGCTGCTTATGAAGCTGCAGACGTTGC
     aPheSerSerLeuAlaGluAlaAlaTyrGluAlaAlaAspValAl
```

```
1801 TGATGATCAGGAAGAACCAGCTACTTACTGCTTATCTTCTTCATT
     aAspAspGlnGluGluProAlaThrTyrCysLeuSerSerSerPh

1846 TGAACTCATAGTTCAGAAGCTCCTAGAGACTACAGACAGACCTGA
     eGluLeuIleValGlnLysLeuLeuGluThrThrAspArgProAs

1891 TGGACACCAGAACAACCTGAGGAGTTCTGCATATGAATCTCTGAT
     pGlyHisGlnAsnAsnLeuArgSerSerAlaTyrGluSerLeuMe

1936 GGAAATTGTGAAAAACAGTGCCAAGGATTGTTACCCTGCTGTCCA
     tGluIleValLysAsnSerAlaLysAspCysTyrProAlaValGl

1981 GAAAACGACTTTGGTCATCATGGAACGACTGCAACAGGTTCTTCA
     nLysThrThrLeuValIleMetGluArgLeuGlnGlnValLeuGl

2026 GATGGAGTCACATATCCAGAGCACATCCGATAGAATCCAGTTCAA
     nMetGluSerHisIleGlnSerThrSerAspArgIleGlnPheAs

2071 TGACCTTCAGTCTTTACTCTGTGCAACTCTTCAGAATGTTCTTCG
     nAspLeuGlnSerLeuLeuCysAlaThrLeuGlnAsnValLeuAr

2116 GAAAGTGCAACATCAAGATGCTTTGCAGATCTCTGATGTGGTTAT
     gLysValGlnHisGlnAspAlaLeuGlnIleSerAspValValMe

2161 GGCCTCCCTGTTAAGGATGTTCCAAAGCACAGCTGGGTCTGGGGG
     tAlaSerLeuLeuArgMetPheGlnSerThrAlaGlySerGlyGl

2206 AGTACAAGAGGATGCCCTGATGGCAGTTAGCACACTGGTGGAAGT
     yValGlnGluAspAlaLeuMetAlaValSerThrLeuValGluVa

2251 GTTGGGTGGTGAATTCCTCAAGTACATGGAGGCCTTTAAACCCTT
     lLeuGlyGlyGluPheLeuLysTyrMetGluAlaPheLysProPh

2296 CCTGGGCATTGGATTAAAAAATTATGCTGAATACCAGGTTTGTTT
     eLeuGlyIleGlyLeuLysAsnTyrAlaGluTyrGlnValCysLe

2341 GGCAGCTGTGGGCTTAGTGGGAGACTTGTGCCGTGCCCTGCAATC
     uAlaAlaValGlyLeuValGlyAspLeuCysArgAlaLeuGlnSe

2386 CAACATCATACCTTTCTGTGACGAGGTGATGCAGCTGCTTCTGGA
     rAsnIleIleProPheCysAspGluValMetGlnLeuLeuLeuGl

2431 AAATTTGGGGAATGAGAACGTCCACAGGTCTGTGAAGCCGCAGAT
     uAsnLeuGlyAsnGluAsnValHisArgSerValLysProGlnIl

2476 TCTGTCAGTGTTTGGTGATATTGCCCTTGCTATTGGAGGAGAGTT
     eLeuSerValPheGlyAspIleAlaLeuAlaIleGlyGlyGluPh

2521 TAAAAAATACTTAGAGGTTGTATTGAATACTCTTCAGCAGGCCTC
     eLysLysTyrLeuGluValValLeuAsnThrLeuGlnGlnAlaSe

2566 CCAAGCCCAGGTGGACAAGTCAGACTATGACATGGTGGATTATCT
     rGlnAlaGlnValAspLysSerAspTyrAspMetValAspTyrLe
```

```
2611 GAATGAGCTAAGGGAAAGCTGCTTGGAAGCCTATACTGGAATCGT
     uAsnGluLeuArgGluSerCysLeuGluAlaTyrThrGlyIleVa

2656 CCAGGGATTAAAGGGGGATCAGGAGAACGTACACCCGGATGTGAT
     lGlnGlyLeuLysGlyAspGlnGluAsnValHisProAspValMe

2701 GCTGGTACAACCCAGAGTAGAATTTATTCTGTCTTTCATTGACCA
     tLeuValGlnProArgValGluPheIleLeuSerPheIleAspHi

2746 CATTGCTGGAGATGAGGATCACACAGATGGAGTAGTAGCTTGTGC
     sIleAlaGlyAspGluAspHisThrAspGlyValValAlaCysAl

2791 TGCTGGACTAATAGGGGACTTATGTACAGCATTTGGGAAGGATGT
     aAlaGlyLeuIleGlyAspLeuCysThrAlaPheGlyLysAspVa

2836 ACTGAAATTAGTAGAAGCTAGGCCAATGATCCATGAATTGTTAAC
     lLeuLysLeuValGluAlaArgProMetIleHisGluLeuLeuTh

2881 TGAAGGGCGGAGATCGAAGACTAACAAAGCAAAAACCCTTGCTAC
     rGluGlyArgArgSerLysThrAsnLysAlaLysThrLeuAlaTh

2926 ATGGGCAACAAAAGAACTGAGGAAACTGAAGAACCAAGCTTGATC
     rTrpAlaThrLysGluLeuArgLysLeuLysAsnGlnAla

2971 TGTTACCATTGGGATGATAACCTGAGGACCCCCACTGGAAATCTC
3016 CCATCTTTTGAAAAACCTGGAAGTGAGGAGTGTGCACGGATGCTG
3061 AATGTTTGGGAATGAGAGGATGAGTGAGTGAGGCTTGAAAACACA
3106 CCACATTGAAAATCCTGCCACAGCAGCAGCCGCAGCCGCCAACAG
3151 CAGCGCTGTTAGTGAGCTAAGTAAGCACTGACTTCGTAGAAAACC
3196 ATAACATCGGCCATCTTGGAAAAGAGAAAAACAATGGAGTTACTT
3241 ATTTAAAAAAAAAGAAAGAAAGTTATCTCTTCCCAGGAGAGGCTA
3286 GAAGTAGCTTTTCTGTCTTTTGGCCAGTGCCGAGTGGAATGCCTG
3331 GTTTCGGGGAGGAGGAGGGACTGGGTTCAGCTGTGGTGCTTTGTT
3376 GTAAAAGGCAGCCTGGCCTTTGCTACTGAGGAGAAAGATGGAGCC
3421 TGGGTCTCAAGCCCACCTTCGCTGTACCTTTGCCACATGGTACTG
3466 TATGCTTGCCAGCTAGAAGGAGGGTCAGGGTTTTTTACAGTCTGA
3511 GAATGAGTGTGTGTGAGTGAGGCGGTATCCACATTCTCAACTTCA
3556 AGTCATTGCTGTTTCTTTTTCCCAGAAAACAAGGGGTTAGATGTT
3601 GCATTTCATAAAACTAACCGAAGTTCTGTCTACTGATGCAGCACA
3646 AGAGATGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACACACA
3691 AAAAAACACACACACAGAGGAAAGACGCTCTTTAGGTTTTGTTTT
3736 GTTTTTTTTTTTTTGGTTTTGTTTTTTGTTTTTTTTTTTACTCTA
3781 GGGAAAACACTGACGAATGGTCAGAGCTCCTATCCTGATCTTTTC
3826 ATCAAGGCGCCTTTCCTAATAATATGGTTCAACTGTGAATGTAGA
3871 AGTGGGGGNNAGGGGGGAGAAAAAGAAAACTCTGGCGTTAGAGGA
3916 TATAAAAAAATATAAGTACAATTGTTACAAATAACGCAGACTTCA
3961 AAAACAAAAAAATCACAACCCAAACAAACCAAAATTTAAATGATC
4006 AGAATTGGCAGCACAAAGAAAACGCCCTCTCCTGACTTGTATTGT
4051 GGCAGTCTGAACGCCCCCAGAAAATTGTGCCAAAGAGTTTAGAAA
4096 AATAAATATACAATAAAAGTAAACACATACACACAAAACAGCAAA
4141 CTTCAGGTAACTATTTTGGATTGCAAACAGGATAATTAAATGTTC
4186 AAACAATCTGATAAAATAACCATTTGG
```

SEQUENCE 8. IMOGEN 38 (Z68747)

```
   1 CGGCGATGTTTCCTAGAGTCTCGACGTTCCTACCTCTTCGCCCCC
         MetPheProArgValSerThrPheLeuProLeuArgProL

  46 TTTCCCGCCACCCTTTGTCCTCTGGAAGCCCGGAGACATCAGCGG
     euSerArgHisProLeuSerSerGlySerProGluThrSerAlaA

  91 CTGCGATTATGCTACTCACTGTTCGGCACGGAACAGTCAGGTACC
     laAlaIleMetLeuLeuThrValArgHisGlyThrValArgTyrA

 136 GCAGTTCAGCGCTGTTGGCCCGGACAAAAAATAACATCCAAAGAT
     rgSerSerAlaLeuLeuAlaArgThrLysAsnAsnIleGlnArgT

 181 ATTTTGGCACTAACAGTGTGATCTGTAGCAAGAAAGATAAGCAGT
     yrPheGlyThrAsnSerValIleCysSerLysLysAspLysGlnS

 226 CTGTTCGAACTGAGGAGACTTCCAAGGAGACTTCAGAGAGCCAAG
     erValArgThrGluGluThrSerLysGluThrSerGluSerGlnA

 271 ACAGTGAAAAGGAAAATACGAAAAAAGACTTGTTAGGCATTATTA
     spSerGluLysGluAsnThrLysLysAspLeuLeuGlyIleIleL

 316 AGGGCATGAAAGTTGAATTAAGCACAGTAAATGTACGAACAACAA
     ysGlyMetLysValGluLeuSerThrValAsnValArgThrThrL

 361 AGCCCCCCAAAAGAAGACCACTTAAAAGTTTGGAAGCTACACTTG
     ysProProLysArgArgProLeuLysSerLeuGluAlaThrLeuG

 406 GCAGGCTTCGAAGAGCTACAGAATATGCTCCAAAGAAGAGAATTG
     lyArgLeuArgArgAlaThrGluTyrAlaProLysLysArgIleG
                                              ▼
 451 AGCCCCTGAGTCCTGAGTTGGTGGCAGCTGCATCTGCTGTGGCAG
     luProLeuSerProGluLeuValAlaAlaAlaSerAlaValAlaA

 496 ATTCTCTCCCTTTTGATAAGCAAACAACCAAGTCAGAGCTGCTGA
     spSerLeuProPheAspLysGlnThrThrLysSerGluLeuLeuS

 541 GCCAGCTCCAGCAGCATGAGGAAGAGTCAAGGGCACAGAGAGATG
     erGlnLeuGlnGlnHisGluGluGluSerArgAlaGlnArgAspA

 586 CAAAGCGACCTAAAATTAGTTTCAGTAACATAATATCAGATATGA
     laLysArgProLysIleSerPheSerAsnIleIleSerAspMetL

 631 AAGTTGCCAGATCTGCTACAGCTAGAGTTCGTTCAAGACCAGAGC
     ysValAlaArgSerAlaThrAlaArgValArgSerArgProGluL

 676 TTCGGATTCAGTTTGATGAAGGCTATGACAATTATCCTGGCCAGG
     euArgIleGlnPheAspGluGlyTyrAspAsnTyrProGlyGlnG

 721 AGAAGACGGATGATCTTAAAAAAAGGAAAAATATATTCACAGGGA
     luLysThrAspAspLeuLysLysArgLysAsnIlePheThrGlyL
```

```
 766 AAAGACTTAATATTTTTGACATGATGGCAGTTACTAAAGAAGCAC
     ysArgLeuAsnIlePheAspMetMetAlaValThrLysGluAlaP

 811 CTGAAACAGACACATCACCTTCACTTTGGAATGTGGAATTTGCTA
     roGluThrAspThrSerProSerLeuTrpAsnValGluPheAlaL

 856 AGCAGTTAGCCACAGTAAATGAACAACCCCTTCAGAATGGATTTG
     ysGlnLeuAlaThrValAsnGluGlnProLeuGlnAsnGlyPheG

 901 AAGAGCTGATCCAGTGGACAAAAGAGGGGAAACTATGGGAGTTCC
     luGluLeuIleGlnTrpThrLysGluGlyLysLeuTrpGluPheP

 946 CAATTAACAATGAAGCAGGTTTTGATGATGATGGTTCAGAATTTC
     roIleAsnAsnGluAlaGlyPheAspAspAspGlySerGluPheH

 991 ATGAACATATATTTCTGGAGAAACACCTGGAGAGCTTTCCAAAAC
     isGluHisIlePheLeuGluLysHisLeuGluSerPheProLysG

1036 AAGGACCAATTCGCCACTTCATGGAGCTGGTGACTTGTGGCCTTT
     lnGlyProIleArgHisPheMetGluLeuValThrCysGlyLeuS

1081 CCAAAAACCCATATCTTAGTGTTAAACAGAAGGTTGAACACATAG
     erLysAsnProTyrLeuSerValLysGlnLysValGluHisIleG

1126 AGTGGTTTAGAAATTATTTTAATGAAAAAAAGGATATTCTAAAAG
     luTrpPheArgAsnTyrPheAsnGluLysLysAspIleLeuLysG

1171 AAAGTAACATACAGTTCAATTAAGACCATGGA
     luSerAsnIleGlnPheAsn
```

SEQUENCE 9. DGS-I (DIGEORGE SYNDROME, L77566)

```
1   GGGGATAGCGATGGAGACGCCGGGCGCATCAGCGTCGTCCTTGTT
     GlyIleAlaMetGluThrProGlyAlaSerAlaSerSerLeuLe

46  GCTTCCCGCCGCGTCCAGGCCCCCGAGGAAGCGCGAGGCGGGAGA
    uLeuProAlaAlaSerArgProProArgLysArgGluAlaGlyGl

91  GGCTGGGGCTGCGACGAGCAAGCAGCGGGTCCTGGACGAGGAAGA
    uAlaGlyAlaAlaThrSerLysGlnArgValLeuAspGluGluGl

136 GTATATCGAGGGCCTCCAGACGGTCATCCAAAGGGATTTCTTTCC
    uTyrIleGluGlyLeuGlnThrValIleGlnArgAspPhePhePr

181 TGATGTGGAGAAGCTCCAGGCACAGAAGGAGTACCTGGAAGCCGA
    oAspValGluLysLeuGlnAlaGlnLysGluTyrLeuGluAlaGl

226 GGAGAATGGAGACTTGGAACGGATGCGCCAGATTGCCATCAAGTT
    uGluAsnGlyAspLeuGluArgMetArgGlnIleAlaIleLysPh

271 TGGCTCTGCCTTGGGCAAGATGTCCCGGGAGCCCCCGCCACCCTA
    eGlySerAlaLeuGlyLysMetSerArgGluProProProProTy

316 TGTGACTCCAGCCACATTTGAAACCCCTGAGGTGCATGCAGGCAC
    rValThrProAlaThrPheGluThrProGluValHisAlaGlyTh

361 TGGAGTGGTGGGCAACAAGCCCAGGCCCCGCGGCCGAGGCCTGGA
    rGlyValValGlyAsnLysProArgProArgGlyArgGlyLeuGl

406 GGATGGAGAGGCTGGAGAGGAGGAGGAGGAAGGAGCCGCTGCCCAG
    uAspGlyGluAlaGlyGluGluGluGluLysGluProLeuProSe

451 CCTAGATGTCTTCCTGAGCCGCTACACGAGTGAGGACAATGCCTC
    rLeu**Asp**ValPheLeuSerArgTyrThrSerGluAspAsnAlaSe

496 CTTCCAGGAGATCATGGAGGTGGCCAAGGAGAGAAGCCGGGCACG
    rPheGlnGluIleMetGluValAlaLysGluArgSerArgAlaAr

541 CCACGCTTGGCTCTACCAGGCTGAGGAAGAGTTTGAGAAGAGGCA
    gHisAlaTrpLeuTyrGlnAlaGluGluGluPheGluLysArgGl

586 GAAAGATAATCTCGAACTCCCGTCAGCAGAGCACCAGGCCATCGA
    nLysAspAsnLeuGluLeuProSerAlaGluHisGlnAlaIleGl

631 GAGCAGCCAGGCCAGTGTGGAGACCTGGAAGTACAAGGCCAAGAA
    uSerSerGlnAlaSerValGluThrTrpLysTyrLysAlaLysAs

676 TTCCCTCATGTACTATCCAGAGGGTGTCCCTGACGAGGAGCAGCT
    nSerLeuMetTyrTyrProGluGlyValProAspGluGluGlnLe

721 GTTTAAGAAGCCCCGGCAGGTGGTACATAAGAACACGCGCTTCCT
    uPheLysLysProArgGlnValValHisLysAsnThrArgPheLe
```

```
 766 TAGGGACCCCTTCAGCCAAGCCCTGAGCAGGTGCCAGCTCCAGCA
     uArgAspProPheSerGlnAlaLeuSerArgCysGlnLeuGlnGl

 811 GGCAGCCGCCCTCAATGCCCAGCACAAACAGGGCAAGGTGGGCCC
     nAlaAlaAlaLeuAsnAlaGlnHisLysGlnGlyLysValGlyPr

 856 CGATGGCAAGGAGCTGATCCCCCAGGAGTCCCCTCGAGTGGGTGG
     oAspGlyLysGluLeuIleProGlnGluSerProArgValGlyGl

 901 ATTTGGATTTGTTGCCACTCCTTCCCCTGCCCCTGGTGTGAACGA
     yPheGlyPheValAlaThrProSerProAlaProGlyValAsnGl

 946 GTCCCCGATGATGACCTGGGGGGAGGTTGAGAACACACCCTTGAG
     uSerProMetMetThrTrpGlyGluValGluAsnThrProLeuAr

 991 AGTTGAAGGGTCGGAAACGCCCTACGTGGACAGGACACCCGGCCC
     gValGluGlySerGluThrProTyrValAspArgThrProGlyPr

1036 AGCTTTTAAGATCCTGGAGCCAGGCCGCAGGGAGCGGCTGGGTCT
     oAlaPheLysIleLeuGluProGlyArgArgGluArgLeuGlyLe

1081 GAAGATGGCCAACGAGGCCGCTGCCAAGAACCGGGCCAAGAAGCA
     uLysMetAlaAsnGluAlaAlaAlaLysAsnArgAlaLysLysGl

1126 GGAAGCCTTGCGGAGAGTGACGGAGAATCTGGCCAGCCTCACCCC
     nGluAlaLeuArgArgValThrGluAsnLeuAlaSerLeuThrPr

1171 CAAAGGCCTGAGCCCAGCCATGTCGCCAGCCCTACAGCGCCTTGT
     oLysGlyLeuSerProAlaMetSerProAlaLeuGlnArgLeuVa

1216 GAGCAGGACGGCCAGCAAGTACACAGACCGGGCCCTGCGGGCCAG
     lSerArgThrAlaSerLysTyrThrAspArgAlaLeuArgAlaSe

1261 CTACACACCATCCCCAGCACGCTCCACCCACCTCAAGACCCCGGC
     rTyrThrProSerProAlaArgSerThrHisLeuLysThrProAl

1306 CAGTGGGCTGCAGACCCCCACAAGCACACCGGCGCCTGGCTCTGC
     aSerGlyLeuGlnThrProThrSerThrProAlaProGlySerAl

1351 CACACGCACCCCTCTCACACAGGACCCGGCCTCCATCACGGACAA
     aThrArgThrProLeuThrGlnAspProAlaSerIleThrAspAs

1396 CCTGCTGCAGCTCCCTGCCCGGCGCAAAGCTTCGGACTTCTTTTA
     nLeuLeuGlnLeuProAlaArgArgLysAlaSerAspPhePhe
1441 GAGCCAGGCCTGGGCTGGGCTCATAGACGCTTCACAGAGCCTGCA
1486 GGGCAGCTGTACACCCAGCAGAGGACTCCAGCCTTCTCGGGGCCC
1531 AGGCCTGGGCCAGAAGCTGTTGACCATACCAGGAGTCACTGGAGA
1576 AAGGGGCTGTGCTGGGGCCAGACTGGCACAAGGCACTCGTGCCCA
1621 CACCACACCCCAGGGCCTTGCCAAGCTGTTTGCTGTTTAATTGGC
1666 CCCTTGAACTGTCATTAAAGAACACCTAGGTAC
```

SEQUENCE 10. ATROPHIN 1 (U23851)

```
   1  TTGGGGTGGAGCAGAGAAGTTTCTGTATTCAGCTGCCCAGGCAGA

  46  GGAGAATGGGGTCTCCACAGCCTGAAGAATGAAGACACGACAGAA
                                          MetLysThrArgGlnAs

  91  TAAAGACTCGATGTCAATGAGGAGTGGACGGAAGAAAGAGGCCCC
      nLysAspSerMetSerMetArgSerGlyArgLysLysGluAlaPr

 136  TGGGCCCCGGGAAGAACTGAGATCGAGGGGCCGGGCCTCCCCTGG
      oGlyProArgGluGluLeuArgSerArgGlyArgAlaSerProGl

 181  AGGGGTCAGCACGTCCAGCAGTGATGGCAAAGCTGAGAAGTCCAG
      yGlyValSerThrSerSerSerAspGlyLysAlaGluLysSerAr

 226  GCAGACAGCCAAGAAGGCCCGAGTAGAGGAAGCCTCCACCCCAAA
      gGlnThrAlaLysLysAlaArgValGluGluAlaSerThrProLy

 271  GGTCAACAAGCAGGGTCGGAGTGAGGAGATCTCAGAGAGTGAAAG
      sValAsnLysGlnGlyArgSerGluGluIleSerGluSerGluSe

 316  TGAGGAGACCAATGCACCAAAAAAGACCAAAACTGAGGAACTCCC
      rGluGluThrAsnAlaProLysLysThrLysThrGluGluLeuPr

 361  TCGGCCACAGTCTCCCTCCGATCTGGATAGCTTGGACGGGCGGAG
      oArgProGlnSerProSerAspLeuAspSerLeuAspGlyArgSe

 406  CCTTAATGATGATGGCAGCAGCGACCCTAGGGATATCGACCAGGA
      rLeuAsnAspAspGlySerSerAspProArgAspIleAspGlnAs

 451  CAACCGAAGCACGTCCCCCAGTATCTACAGCCCTGGAAGTGTGGA
      pAsnArgSerThrSerProSerIleTyrSerProGlySerValGl

 496  GAATGACTCTGACTCATCTTCTGGCCTGTCCCAGGGCCCAGCCCG
      uAsnAspSerAspSerSerSerGlyLeuSerGlnGlyProAlaAr

 541  CCCCTACCACCCACCTCCACTCTTTCCTCCTTCCCCTCAACCGCC
      gProTyrHisProProProLeuPheProProSerProGlnProPr

 586  AGACAGCACCCCTCGACAGCCAGAGGCTAGCTTTGAACCCCATCC
      oAspSerThrProArgGlnProGluAlaSerPheGluProHisPr

 631  TTCTGTGACACCCACTGGATATCATGCTCCCATGGAGCCCCCCAC
      oSerValThrProThrGlyTyrHisAlaProMetGluProProTh

 676  ATCTCGAATGTTCCAGGCTCCTCCTGGGGCCCCTCCCCCTCACCC
      rSerArgMetPheGlnAlaProProGlyAlaProProProHisPr

 721  ACAGCTCTATCCCGGGGGCACTGGTGGAGTTTTGTCTGGACCCCC
      oGlnLeuTyrProGlyGlyThrGlyGlyValLeuSerGlyProPr
```

```
766  AATGGGTCCCAAGGGGGGAGGGGCTGCCTCATCAGTGGGGGGCCC
     oMetGlyProLysGlyGlyGlyAlaAlaSerSerValGlyGlyPr

811  TAATGGGGGTAAGCAGCACCCCCCACCCACTACTCCCATTTCAGT
     oAsnGlyGlyLysGlnHisProProProThrThrProIleSerVa

856  ATCAAGCTCTGGGGCTAGTGGTGCTCCCCCAACAAAGCCGCCTAC
     lSerSerSerGlyAlaSerGlyAlaProProThrLysProProTh

901  CACTCCAGTGGGTGGTGGGAACCTACCTTCTGCTCCACCACCAGC
     rThrProValGlyGlyGlyAsnLeuProSerAlaProProProAl

946  CAACTTCCCCCATGTGACACCGAACCTGCCTCCCCCACCTGCCCT
     aAsnPheProHisValThrProAsnLeuProProProProAlaLe

991  GAGACCCCTCAACAATGCATCAGCCTCTCCCCCTGGCCTGGGGGC
     uArgProLeuAsnAsnAlaSerAlaSerProProGlyLeuGlyAl

1036 CCAACCACTACCTGGTCATCTGCCCTCTCCCCACGCCATGGGACA
     aGlnProLeuProGlyHisLeuProSerProHisAlaMetGlyGl

1081 GGGTATCGGTGGACTTCCTCCTGGCCCAGAGAAGGGCCCAACTCT
     nGlyIleGlyGlyLeuProProGlyProGluLysGlyProThrLe

1126 GGCTCCTTCACCCCACTCTCTGCCTCCTGCTTCCTCTTCTGCTCC
     uAlaProSerProHisSerLeuProProAlaSerSerSerAlaPr

1171 AGCGCCCCCCATGAGGTTTCCTTATTCATCCTCTAGTAGTAGCTC
     oAlaProProMetArgPheProTyrSerSerSerSerSerSerSe

1216 TGCAGCAGCCTCCTCTTCCAGTTCTTCCTCCTCTTCCTCTGCCTC
     rAlaAlaAlaSerSerSerSerSerSerSerSerSerSerAlaSe

1261 CCCCTTCCCAGCTTCCCAGGCATTGCCCAGCTACCCCCACTCTTT
     rProPheProAlaSerGlnAlaLeuProSerTyrProHisSerPh

1306 CCCTCCCCCAACAAGCCTCTCTGTCTCCAATCAGCCCCCCAAGTA
     eProProProThrSerLeuSerValSerAsnGlnProProLysTy

1351 TACTCAGCCTTCTCTCCCATCCCAGGCTGTGTGGAGCCAGGGTCC
     rThrGlnProSerLeuProSerGlnAlaValTrpSerGlnGlyPr

1396 CCCACCACCTCCTCCCTATGGCCGCCTCTTAGCCAACAGCAATGC
     oProProProProProTyrGlyArgLeuLeuAlaAsnSerAsnAl

1441 CCATCCAGGCCCCTTCCCTCCCTCTACTGGGGCCCAGTCCACCGC
     aHisProGlyProPheProProSerThrGlyAlaGlnSerThrAl

1486 CCACCCACCAGTCTCAACACATCACCATCACCACCAGCAACAGCA
     aHisProProValSerThrHisHisHisHisGlnGlnGlnGl

1531 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCATCACGGAAACTC
     nGlnGlnGlnGlnGlnGlnGlnGlnGlnHisHisGlyAsnSe
```

```
1576  TGGGCCCCCTCCTCCTGGAGCATTTCCCCACCCACTGGAGGGCGG
      rGlyProProProProGlyAlaPheProHisProLeuGluGlyGl

1621  TAGCTCCCACCACGCACACCCTTACGCCATGTCTCCCTCCCTGGG
      ySerSerHisHisAlaHisProTyrAlaMetSerProSerLeuGl

1666  GTCTCTGAGGCCCTACCCACCAGGGCCAGCACACCTGCCCCCACC
      ySerLeuArgProTyrProProGlyProAlaHisLeuProProPr

1711  TCACAGCCAGGTGTCCTACAGCCAAGCAGGCCCCAATGGCCCTCC
      oHisSerGlnValSerTyrSerGlnAlaGlyProAsnGlyProPr

1756  AGTCTCTTCCTCTTCCAACTCTTCCTCTTCCACTTCTCAAGGGTC
      oValSerSerSerSerAsnSerSerSerSerThrSerGlnGlySe

1801  CTACCCATGTTCACACCCCTCCCCTTCCCAGGGCCCTCAAGGGGC
      rTyrProCysSerHisProSerProSerGlnGlyProGlnGlyAl

1846  GCCCTACCCTTTCCCACCGGTGCCTACGGTCACCACCTCTTCGGC
      aProTyrProPheProProValProThrValThrThrSerSerAl

1891  TACCCTTTCCACGGTCATTGCCACCGTGGCTTCCTCGCCAGCAGG
      aThrLeuSerThrValIleAlaThrValAlaSerSerProAlaGl

1936  CTACAAAACGGCCTCCCCACCTGGGCCCCCACCGTACGGAAAGAG
      yTyrLysThrAlaSerProProGlyProProProTyrGlyLysAr

1981  AGCCCCGTCCCCGGGGGCCTACAAGACAGCCACCCCACCCGGATA
      gAlaProSerProGlyAlaTyrLysThrAlaThrProProGlyTy

2026  CAAACCCGGGTCGCCTCCCTCCTTCCGAACGGGGACCCCACCGGG
      rLysProGlySerProProSerPheArgThrGlyThrProProGl

2071  CTATCGAGGAACCTCGCCACCTGCAGGCCCAGGGACCTTCAAGCC
      yTyrArgGlyThrSerProProAlaGlyProGlyThrPheLysPr

2116  GGGCTCGCCCACCGTGGGACCTGGGCCCCTGCCACCTGCGGGGCC
      oGlySerProThrValGlyProGlyProLeuProProAlaGlyPr

2161  CTCAGGCCTGCCATCGCTGCCACCACCACCTGCGGCCCCTGCCTC
      oSerGlyLeuProSerLeuProProProProAlaAlaProAlaSe

2206  AGGGCCGCCCCTGAGCGCCACGCAGATCAAACAGGAGCCGGCTGA
      rGlyProProLeuSerAlaThrGlnIleLysGlnGluProAlaGl

2251  GGAGTATGAGACCCCCGAGAGCCCGGTGCCCCCAGCCCGCAGCCC
      uGluTyrGluThrProGluSerProValProProAlaArgSerPr

2296  CTCGCCCCCTCCCAAGGTGGTAGATGTACCCAGCCATGCCAGTCA
      oSerProProProLysValValAspValProSerHisAlaSerGl

2341  GTCTGCCAGGTTCAACAAACACCTGGATCGCGGCTTCAACTCGTG
      nSerAlaArgPheAsnLysHisLeuAspArgGlyPheAsnSerCy
```

```
2386  CGCGCGCAGCGACCTGTACTTCGTGCCACTGGAGGGCTCCAAGCT
      sAlaArgSerAspLeuTyrPheValProLeuGluGlySerLysLe

2431  GGCCAAGAAGCGGGCCGACCTGGTGGAGAAGGTGCGGCGCGAGGC
      uAlaLysLysArgAlaAspLeuValGluLysValArgArgGluAl

2476  CGAGCAGCGCGCGCGCGAAGAAAAGGAGCGCGAGCGCGAGCGGGA
      aGluGlnArgAlaArgGluGluLysGluArgGluArgGluArgGl

2521  ACGCGAGAAAGAGCGCGAGCGCGAGAAGGAGCGCGAGCTTGAACG
      uArgGluLysGluArgGluArgGluLysGluArgGluLeuGluAr

2566  CAGCGTGAAGTTGGCTCAGGAGGGCCGTGCTCCGGTGGAATGCCC
      gSerValLysLeuAlaGlnGluGlyArgAlaProValGluCysPr

2611  ATCTCTGGGCCCAGTGCCCCATCGCCCTCCATTTGAACCGGGCAG
      oSerLeuGlyProValProHisArgProProPheGluProGlySe
                    ▼
2656  TGCGGTGGCTACAGTGCCCCCCTACCTGGGTCCTGACACTCCAGC
      rAla**Val**AlaThrValProProTyrLeuGlyProAspThrProAl

2701  CTTGCGCACTCTCAGTGAATATGCCCGGCCTCATGTCATGTCTCC
      aLeuArgThrLeuSerGluTyrAlaArgProHisValMetSerPr

2746  TGGCAATCGCAACCATCCATTCTACGTGCCCCTGGGGGCAGTGGA
      oGlyAsnArgAsnHisProPheTyrValProLeuGlyAlaValAs

2791  CCCGGGGCTCCTGGGTTACAATGTCCCGGCCCTGTACAGCAGTGA
      pProGlyLeuLeuGlyTyrAsnValProAlaLeuTyrSerSerAs

2836  TCCAGCTGCCCGGGAGAGGGAACGGGAAGCCCGTGAACGAGACCT
      pProAlaAlaArgGluArgGluArgGluAlaArgGluArgAspLe

2881  CCGTGACCGCCTCAAGCCTGGCTTTGAGGTGAAGCCTAGTGAGCT
      uArgAspArgLeuLysProGlyPheGluValLysProSerGluLe

2926  GGAACCCCTACATGGGGTCCCTGGGCCGGGCTTGGATCCCTTTCC
      uGluProLeuHisGlyValProGlyProGlyLeuAspProPhePr

2971  CCGACATGGGGGCCTGGCTCTGCAGCCTGGCCCACCTGGCCTGCA
      oArgHisGlyGlyLeuAlaLeuGlnProGlyProProGlyLeuHi

3016  CCCTTTCCCCTTTCATCCGAGCCTGGGGCCCCTGGAGCGAGAACG
      sProPheProPheHisProSerLeuGlyProLeuGluArgGluAr

3061  TCTAGCGCTGGCAGCTGGGCCAGCCCTGCGGCCTGACATGTCCTA
      gLeuAlaLeuAlaAlaGlyProAlaLeuArgProAspMetSerTy

3106  TGCTGAGCGGCTGGCAGCTGAGAGGCAGCACGCAGAAAGGGTGGC
      rAlaGluArgLeuAlaAlaGluArgGlnHisAlaGluArgValAl

3151  GGCCCTGGGCAATGACCCACTGGCCCGGCTGCAGATGCTCAATGT
      aAlaLeuGlyAsnAspProLeuAlaArgLeuGlnMetLeuAsnVa
```

```
3196  GACTCCCCATCACCACCAGCACTCCCACATCCACTCGCACCTGCA
      lThrProHisHisHisGlnHisSerHisIleHisSerHisLeuHi

3241  CCTGCACCAGCAAGATGCTATCCATGCAGCCTCTGCCTCGGTGCA
      sLeuHisGlnGlnAspAlaIleHisAlaAlaSerAlaSerValHi

3286  CCCTCTCATTGACCCCCTGGCCTCAGGGTCTCACCTTACCCGGAT
      sProLeuIleAspProLeuAlaSerGlySerHisLeuThrArgIl

3331  CCCCTACCCAGCTGGAACTCTCCCTAACCCCCTGCTTCCTCACCC
      eProTyrProAlaGlyThrLeuProAsnProLeuLeuProHisPr

3376  TCTGCACGAGAACGAAGTTCTTCGTCACCAGCTCTTTGCTGCCCC
      oLeuHisGluAsnGluValLeuArgHisGlnLeuPheAlaAlaPr

3421  TTACCGGGACCTGCCGGCCTCCCTTTCTGCCCCGATGTCAGCAGC
      oTyrArgAspLeuProAlaSerLeuSerAlaProMetSerAlaAl

3466  TCATCAGCTGCAGGCCATGCACGCACAGTCAGCTGAGCTGCAGCG
      aHisGlnLeuGlnAlaMetHisAlaGlnSerAlaGluLeuGlnAr

3511  CTTGGCGCTGGAACAGCAGCAGTGGCTGCATGCCCATCACCCGCT
      gLeuAlaLeuGluGlnGlnGlnTrpLeuHisAlaHisHisProLe

3556  GCACAGTGTGCCGCTGCCTGCCCAGGAGGACTACTACAGTCACCT
      uHisSerValProLeuProAlaGlnGluAspTyrTyrSerHisLe

3601  GAAGAAGGAAAGCGACAAGCCACTGTAGAACCTGCGATCAAGAGA
      uLysLysGluSerAspLysProLeu

3646  GCACCATGGCTCCTACATTGGACCTTGGAGCACCCCCACCCTCCC
3691  CCCACCGTGCCCTTGGCCTGCCACCCAGAGCCAAGAGGGTGCTGC
3736  TCAGTTGCAGGGCCTCCGCAGCTGGACAGAGAGTGGGGGAGGGAG
3781  GGACAGACAGAAGGCCAAGGCCCGATGTGGTGTGCAGAGGTGGGG
3826  AGGTGGCGAGGATGGGGACAGAAAGCGCACAGAATCTTGGACCAG
3871  GTCTCTCTTCCTTGTCCCCCCTGCTTTTCTCCTCCCCCATGCCCA
3916  ACCCCTGTGGCCGCCGCCCCTCCCCTGCCCCGTTGGTGTGATTAT
3961  TTCATCTGTTAGATGTGGCTGTTTTGCGTAGCATCGTGTGCCACC
4006  CCTGCCCCTCCCCGATCCCTGTGTGCGCGCCCCCTCTGCAATGTA
4051  TGCCCCTTGCCCCTTCCCCACACTAATAATTTATATATATAAATA
4096  TCTATATGACGCTCTTAAAAAAACATCCCAACCAAAACCAACCAA
4141  ACAAAAACATCCTCACAACTCCCCAGGA
```

SEQUENCE 11. EMBRYONAL FYN SUBSTRATE (EFS2, AB001467)

```
   1  CTCCAGGCAACTTGGGGCAAGCGTCTCAGTTCTCGCTCTCCCTTC
  46  CTCCCAGCGGGGTCGCCGCAGACCCCAGCCCTGGGAGCACCGCTC
  91  TGCAGCGCGGCCGGCGGGTGGAGACGGTTGGCCCCTAAACTCGCT
 136  CGTCCAGCCCAACCGCCCCGGCGGCTTCTCCCAGCCCTCGAGGCT
 181  CTCCTGAGCGGCCTGGAGAGGCGTCGAGCGCAGCCCAGCGCCCGC
 226  CTGCTCACCCGCCCCGGCCCGGGAAGGGAATTTTCGGATCCTGCG
 271  AGCCCGGGGCGCCCCCGCGGCCTAGGGCGGGCAGCTCCCGGGGCC
 316  TGGCCGAGCCGGTGGCGCCCGGGAGGCCGCGGGGACAGCACGCAG
 361  CGCGCGCCCTTGGATGCCGTCCCGCAGCGACGCCCCGGCCCGCCC
 406  CGCTCCTCCTCCTGCCTGGCTAGCCTGCCTCTCATTTGGGAAGTT
 451  TTGTGGGTTTTCTTTCTCCTCCTCCAACCTTGGCGGAGGCCACGA
 496  CTCAGGCGCCACAGCTGGGGGCTAGAGGCCGCGGACCATGGTGCG
 541  GGGCAGCCACCGCTGAAGTCAGCAAAACCGAGCCTGGCCTGAGGC

 586  AGGCTGCGCGGGAGGCCAAAGCCATGGCCATTGCCACGTCGACCC
                                MetAlaIleAlaThrSerThrG

 631  AGCTGGCCCGGGCACTGTATGACAACACCGCTGAGTCCCCCCAGG
      lnLeuAlaArgAlaLeuTyrAspAsnThrAlaGluSerProGlnG

 676  AGCTGTCCTTCCGCCGAGGGGATGTCCTACGGGTCCTGCAGAGAG
      luLeuSerPheArgArgGlyAspValLeuArgValLeuGlnArgG

 721  AGGGCGCTGGTGGACTGGACGGCTGGTGCCTCTGCTCCCTACACG
      luGlyAlaGlyGlyLeuAspGlyTrpCysLeuCysSerLeuHisG

 766  GCCAGCAGGGCATTGTGCCCGCCAACAGGGTGAAGCTCTTGCCTG
      lyGlnGlnGlyIleValProAlaAsnArgValLysLeuLeuProA

 811  CTGGCCCAGCACCCAAGCCCAGCCTCTCTCCTGCGTCCCCAGCCC
      laGlyProAlaProLysProSerLeuSerProAlaSerProAlaG

 856  AGCCTGGCTCACCATATCCAGCCCCAGATCACAGCAATGAGGACC
      lnProGlySerProTyrProAlaProAspHisSerAsnGluAspG

 901  AGGAGGTGTATGTGGTGCCGCCCCCAGCTCGGCCCTGTCCAACCT
      lnGluValTyrValValProProProAlaArgProCysProThrS

 946  CAGGACCTCCAGCTGGACCTTGCCCACCCTCTCCTGACCTCATCT
      erGlyProProAlaGlyProCysProProSerProAspLeuIleT

 991  ACAAAATCCCCAGAGCTAGTGGGACCCAGCTGGCTGCTCCCAGAG
      yrLysIleProArgAlaSerGlyThrGlnLeuAlaAlaProArgA

1036  ATGCCTTGGAGGTCTACGATGTGCCCCCCACCGCCCTCCGGGTGC
      spAlaLeuGluValTyrAspValProProThrAlaLeuArgValP

1081  CCTCCAGTGGCCCCTATGACTGCCCTGCCTCCTTTTCCCACCCTC
      roSerSerGlyProTyrAspCysProAlaSerPheSerHisProL
```

```
1126  TGACCCGGGTTGCCCCGCAGCCCCCTGGAGAGGATGATGCTCCCT
      euThrArgValAlaProGlnProProGlyGluAspAspAlaProT

1171  ATGATGTGCCTCTGACCCCAAAGCCACCTGCAGAGCTGGAACCAG
      yrAspValProLeuThrProLysProProAlaGluLeuGluProA

1216  ATCTGGAGTGGGAAGGAGGCCGGGAGCCGGGGCCCCCCATCTATG
      spLeuGluTrpGluGlyGlyArgGluProGlyProProIleTyrA

1261  CTGCCCCCTCCAACCTGAAACGAGCGTCAGCCTTACTCAATTTGT
      laAlaProSerAsnLeuLysArgAlaSerAlaLeuLeuAsnLeuT

1306  ATGAAGCACCCGAGGAACTGCTGGCAGACGGGGAGGGCGGGGGCA
      yrGluAlaProGluGluLeuLeuAlaAspGlyGluGlyGlyGlyT

1351  CTGATGAGGGGATCTACGATGTGCCTCTGCTGGGGCCAGAGGCTC
      hrAspGluGlyIleTyrAspValProLeuLeuGlyProGluAlaP

1396  CCCCTTCTCCAGAGCCCCCTGGAGCCTTGGCCTCCCATGACCAGG
      roProSerProGluProProGlyAlaLeuAlaSerHisAspGlnA

1441  ACACCCTGGCCCAGCTTCTGGCCAGAAGCCCCCCACCCCCACACA
      spThrLeuAlaGlnLeuLeuAlaArgSerProProProProHisA

1486  GGCCCCGGCTCCCCTCAGCTGAGAGCCTGTCCCGCCGCCCTCTGC
      rgProArgLeuProSerAlaGluSerLeuSerArgArgProLeuP

1531  CTGCCCTGCCTGTCCCTGAGGCCCCCAGCCCCTCCCCAGTGCCCT
      roAlaLeuProValProGluAlaProSerProSerProValProS

1576  CTCCTGCCCCAGGCCGGAAGGGCAGCATCCAGGACCGGCCTCTGC
      erProAlaProGlyArgLysGlySerIleGlnAspArgProLeuP
                                                  ▼
1621  CCCCACCCCCACCCCGCCTGCCTGGTTATGGAGGCCCCAAGGTCG
      roProProProProArgLeuProGlyTyrGlyGlyProLysValG

1666  AGGGGGATCCAGAGGGCAGGGAGATGGAGGATGACCCAGCAGGAC
      luGlyAspProGluGlyArgGluMetGluAspAspProAlaGlyH

1711  ACCACAATGAGTACGAGGGCATTCCGATGGCCGAGGAGTATGACT
      isHisAsnGluTyrGluGlyIleProMetAlaGluGluTyrAspT

1756  ATGTCCACCTGAAGGGCATGGACAAAGCTCAGGGATCTAGGCCCC
      yrValHisLeuLysGlyMetAspLysAlaGlnGlySerArgProP

1801  CGGATCAGGCCTGCACAGGGGATCCTGAACTGCCCGAGAGGGGGA
      roAspGlnAlaCysThrGlyAspProGluLeuProGluArgGlyM

1846  TGCCGGCGCCGCAGGAGGCCCTGTCCCCAGGGGAGCCACTGGTTG
      etProAlaProGlnGluAlaLeuSerProGlyGluProLeuValV

1891  TGTCCACCGGAGATCTGCAGCTCCTGTACTTCTATGCTGGGCAAT
      alSerThrGlyAspLeuGlnLeuLeuTyrPheTyrAlaGlyGlnC
```

```
1936  GCCAGAGCCACTACTCAGCCCTGCAGGCAGCCGTGGCAGCCCTGA
      ysGlnSerHisTyrSerAlaLeuGlnAlaAlaValAlaAlaLeuM

1981  TGTCCAGTACCCAGGCTAATCAGCCCCCGCGCCTTTTCGTGCCCC
      etSerSerThrGlnAlaAsnGlnProProArgLeuPheValProH

2026  ACAGCAAGAGGGTGGTGGTGGCTGCTCATCGCCTGGTGTTTGTTG
      isSerLysArgValValValAlaAlaHisArgLeuValPheValG

2071  GGGACACCCTGGGCCGGCTGGCAGCCTCTGCCCCTCTGAGAGCAC
      lyAspThrLeuGlyArgLeuAlaAlaSerAlaProLeuArgAlaG

2116  AGGTCAGGGCTGCAGGTACAGCACTGGGCCAGGCATTGCGGGCCA
      lnValArgAlaAlaGlyThrAlaLeuGlyGlnAlaLeuArgAlaT

2161  CTGTGCTGGCTGTCAAGGGAGCTGCCCTGGGCTACCCATCCAGCC
      hrValLeuAlaValLysGlyAlaAlaLeuGlyTyrProSerSerP

2206  CTGCCATCCAAGAGATGGTGCAGTGTGTAACAGAACTGGCAGGGC
      roAlaIleGlnGluMetValGlnCysValThrGluLeuAlaGlyG

2251  AGGCCCTGCAATTCACTACCCTGCTCACTAGCCTGGCTCCATGAA
      lnAlaLeuGlnPheThrThrLeuLeuThrSerLeuAlaPro

2296  GGTCCTTTGGCACAGCTCTGCTCCTCCCCTGCCTGCCAAAGCCCC
2341  CCTTTAGGCCTTGGGTGGCTGGAAGGCTTTGTTAAGGGACTAGGA
2386  GAAATGGGGGTATCTTTCCCCTTTCCTGCCCTTTCTGCTCATCTC
2431  AACCTCTCACAGAGGTGTCTTCTCCCCCTAACCTACAGCTTTTTG
2476  TACAAGCCATTTTGTGTAAATTATTTATATTTAATATTATTCCCT
2521  GCTTTGTCAGGAGCAGGTACTAGGCTCTGGGGCAGTGAGGAACTA
2566  GATCCTTCTCTCCTCAGCCTAGGGTGGAGGTCACTGCACTACCAC
2611  CCACCTCTGGAAGACTGGCTGTGAAAAGTCAGGTGGCAGAAACCT
2656  GGGGCCACATAGAGCCTCTCTCTTTTCCTGTTTCTTGGCTCTAGA
2701  AGATCAGCACTGCACTGTTAGCTGAGAGTGCGGGCAAGACATAAA
2746  CTGTCCAGAGTTTGAAGGTTCTCGGAAAGACCGGAGGGCTTCTCC
2791  CCACAGAAGGCGGAGAGAGCTGGGGCTCAGACATGGGTGTGCACC
2836  TTAATAAACCCTGCTGTCTGCCTCCCTGACTCTGCTTCTTGGGAG
2881  CATGGTGAGCAGCCCTGGTGCTCAGCAGCCATACCTATGGGACAC
2926  ACACTACGAAAAGGATGCCTTTAGGGTTTGGGGGAGATTTTACTC
2971  CTTTCTTCAACAACTATTCACTGGACAAGTTCTCTGCTCCCATGA
3016  CGCGCCAGGCACAGTTCTGCAAGTATATTGTGAATGTATTGTTCT
3061  AGTGGGATACACAAATAAGTCAGTTAAAATACATAAATAAAAACA
3106  TAAACCTGC
```

SEQUENCE 12. HUMAN BRAIN FACTOR 2 (HBF-G2, X78202)

```
   1  GGCGAGAGAGACGCTCCCGCTCGCCGCCAGCTCTGATTGGCCCAG
  46  CGGTAGGAAAGGTTAAACCAAAAATTTTTTTACAGCCCTAGTGTG
  91  CGCCTGTAGCTCGGAAAATTAATTGTGGCTATAGCCGCCTCGATC
 136  GCTGTCTCCCCAGCCTCGCCGCGGACGCTCCGGGACGCGCCCCGCC
 181  CGCCGCCCGGTTCTCCCCCCCTTTGGGCTGGTGCTGCTGCTGCTG
 226  TGACTGCTGCTGCGAAAGGAGGAGGAGGAGGAGGAAGCAGCGGGG
 271  GGGGGAGCGGTGGGTGTGGGGGAAACCAAGAGTACAGTGGACGAG
 316  GACTCACCCCGGCGTGGTGTTCTTTTTTCTTCTTCTTTTTCTTTC
 361  CTTTTTTTTTTTTTTTTCTAATTCCTGAGGGGTGGTTGCTGCTTT
 406  TGCTACATGACTTGCCAGCGCCCGAGCCTGCGGTCCAACTGCGCT
 451  GCTGCCGGAGCGCTCAGTGCCGCCGCTGCCGCCCGTGCCCCCCGC
 496  GCCCCGTTCGGCACCCACCGGTCGCCGCCCCGCCCGCGCGCCGCT
 541  GTCCCGCTCCCGCGCCGCCGCCGCCGTTTCCCCCCGACGACTGGG

 586  TGATGCTGGACATGGGAGATAGGAAAGAGGTGAAAATGATCCCCA
        MetLeuAspMetGlyAspArgLysGluValLysMetIleProL

 631  AGTCCTCGTTCAGCATCAACAGCCTGGTGCCCGAGGGCCTCCAGA
      ysSerSerPheSerIleAsnSerLeuValProGluGlyLeuGlnA

 676  ACGACAACCACCACGCGAGCCACGGCCACCACAACAGCCACCACC
      snAspAsnHisHisAlaSerHisGlyHisHisAsnSerHisHisP

 721  CCCAGCACCACCACCACCACCACCATCACCACCACCCGCCGC
      roGlnHisHisHisHisHisHisHisHisHisProProP

 766  CGCCCGCCCCGCAACCGCCGCCGCCGCCGCAGCAGCAGCAGCGC
      roProAlaProGlnProProProProGlnGlnGlnGlnProP

 811  CGCCGCCGCCGAGACGCGGGGCCCGGCGCCGACGACGACGAGGCC
      roProProProArgArgGlyAlaArgArgArgArgArgGlyP

 856  CCAGCAGTTGTTGTTCCGCCGCCACGCACACGGCGCGCCTGAGG
      roSerSerCysCysSerAlaAlaHisAlaHisGlyAlaProGluG

 901  GCCAACGGCAGCTGGCGCAAGGCGACCGGCGCGGCCGGGGGATCT
      lyGlnArgGlnLeuAlaGlnGlyAspArgArgGlyArgGlyIleC

 946  GCCCCGTCGGGCCGGACGAGAAGGAGAAGGCCCCGCGCCGGGGGGG
      ysProValGlyProAspGluLysGluLysAlaArgAlaGlyGlyG

 991  AGGAGAAGAAGGGGGCGGGCGAGGGCGGCAAGGACGGGGAGGGGG
      luGluLysLysGlyAlaGlyGluGlyGlyLysAspGlyGluGlyG

1036  GCAAGGAGGGCGAGAAGAAGAACGGCAAGTACGAGAAGCCGCCGT
      lyLysGluGlyGluLysLysAsnGlyLysTyrGluLysProProP

1081  TCAGCTACAACGCGCTCATCATGATGGCCATGCGGCAGAGCCCCG
      heSerTyrAsnAlaLeuIleMetMetAlaMetArgGlnSerProG
```

```
1126  AGAAGCGGCTCACGCTCAACGGCATCTACGAGTTCATCATGAAGA
      luLysArgLeuThrLeuAsnGlyIleTyrGluPheIleMetLysA

1171  ACTTCCCTTACTACCGCGAGAACAAGCAGGGCTGGCAGAACTCCA
      snPheProTyrTyrArgGluAsnLysGlnGlyTrpGlnAsnSerI

1216  TCCGCCACAATCTGTCCCTCAACAAGTGCTTCGTGAAGGTGCCGC
      leArgHisAsnLeuSerLeuAsnLysCysPheValLysValProA

1261  GCCACTACGACGACCCGGGCAAGGGCAACTACTGGATGCTGGACC
      rgHisTyrAspAspProGlyLysGlyAsnTyrTrpMetLeuAspP

1306  CGTCGAGCGACGACGTGTTCATCGGCGGCACCACGGGCAAGCTGC
      roSerSerAspAspValPheIleGlyGlyThrThrGlyLysLeuA

1351  GGCGCTCCACCACCTCGCCGGCCAAGCCGGCCTTCAAGCGCGGTG
      rgArgSerThrThrSerProAlaLysProAlaPheLysArgGlyA

1396  CCGCGCTCACCTCCACCGGCCTCACCTTCATGGACGCGCCGGCTC
      laAlaLeuThrSerThrGlyLeuThrPheMetAspAlaProAlaP

1441  CCTCTACTGGCCCATGTCGCCCTTCCTGTCCCTGCACCACCCCCG
      roSerThrGlyProCysArgProSerCysProCysThrThrProA

1486  CCAGCAGCACTTTGAGTTACAACGGGACCACGTCGGCCTACCCCA
      laSerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProS

1531  GCCACCCCATGCCCTACAGCTCCGTGTTGACTCAAAACTCGCTGG
      erHisProMetProTyrSerSerValLeuThrGlnAsnSerLeuG

1576  GCAACAACCACTCCTCCTCCACCGCCAACGGGCTGAGCGTGGACC
      lyAsnAsnHisSerSerSerThrAlaAsnGlyLeuSerValAspA
                          ▼
1621  GGCTGGTCAACGGGGGAATCCCGTACGCCACGCACCACCTCACGG
      rgLeuValAsnGlyGly**Ile**ProTyrAlaThrHisHisLeuThrA

1666  CCGCCGCGCTAACCGCCTCGGTGCCCTGCGGCCTGCTGGTGCCCT
      laAlaAlaLeuThrAlaSerValProCysGlyLeuLeuValProC

1711  GCTCTGGGACCTACTCCCTCAACCCCTGCTCCGTCAACCTGCTCG
      ysSerGlyThrTyrSerLeuAsnProCysSerValAsnLeuLeuA

1756  CGGGCCAGACCAGTTACTTTTTCCCCCACGTCCCGCACCCGTCAA
      laGlyGlnThrSerTyrPhePheProHisValProHisProSerM

1801  TGACTTCGCAGAGCAGCACGTCCATGAGCGCCAGGGCCGCGTCCT
      etThrSerGlnSerSerThrSerMetSerAlaArgAlaAlaSerS

1846  CCTCCACGTCGCCGGCAGGCCCCCCTCGACCCCTGCCCTGTGAGT
      erSerThrSerProAlaGlyProProArgProLeuProCysGluS

1891  CTTTAAGACCCTCTTTGCCAAGTTTTACGACGGGACTGTCTGGGG
      erLeuArgProSerLeuProSerPheThrThrGlyLeuSerGlyG
```

```
1936  GACTGTCTGATTATTTCACACATCAAAATCAGGGGTCTTCTTCCA
      lyLeuSerAspTyrPheThrHisGlnAsnGlnGlySerSerSerA

1981  ACCCTTTAATACATTAACATCCCTGGGACCAGACTGTAAGTGAAC
      snProLeuIleHis

2026  GTTTTACACACATTTGCATTGTAAATGATAATTAAAAAAATAAGT
2071  CCAGGTATTTTTTATTAAGCCCCCCCCTCCCATTTCTGTACGTTT
2116  GTTCAGTCTCTAGGGTTGTTTATTATTCTAACAAGGTGTGGAGTG
2161  TCAGCGAGGTGCAATGTGGGGAGAATACATTGTAGAATATAAGGT
2206  TTGGAAGTCAAATTATAGTAGAATGTGTATCTAAATAGTGACTGC
2251  TTTGCCATTTCATTCAAACCTGACAAGTCTATCTCTAAGAGCCGC
2296  CAGATTTCCATGTGTGCAGTATTATAAGTTATCATGGAACTATAT
2341  GGTGGACGCAGACCTTGAGAACAACCTAAATTATGGGGAGAATTT
2386  TAAAATGTTAAACTGTAATTTGTATTTAAAAAGCATTCGTAGTAA
2431  AGGTGCCCAAGAAATTATTTTGGCCATTTATTGTTTTCTCCTTTT
2476  CTTTAAAGAACTGTTTTTTTTTCTTTTGTTTACTTTTAGACCAAA
2521  GATTGGGCGGTTCTAGAAAATGCGCCTTGGTATACTAAGTATTAA
2566  AACAAACAAAAAGGAAAGTTGTTTCAGTTAACGCTGCCCATTCAA
2611  TTGAATCAGAAGGGGACAAAATTAACGATTGCCTTCAGTTTGTGT
2656  TGTGTATATTTTGATGTATGTGGTCACTAACAGGTCACTTTTATT
2701  TTTTCTAAATGTAGTGAAATGTTAATACCTATTGTACTTATAGGT
2746  AAACCTTGCAAATATGTAACCTGTGTTGCGCAAATGCCGCATAAA
2791  TTTGAGTGATTGTTAATGTTGTCTTAAAATTTCTTGATTGTGACT
2836  ATGTGGTCATATGCCCGTGTTTGTCACTTACAAAAATGTTTACTA
2881  TGAACACACATAAATAAAAAATAG
```

SEQUENCE 13. IP-1


```
  1 GAATTCGGACGAGGGAGGAGGAAGACCAAGACATACAGGGAGAAA
    IleArgThrArgGluGluGluAspGlnAspIleGlnGlyGluI

 46 TCAGTCATCCTGATGGAAAGGTGGAAAAGGTTTATAAGAATGGGT
    leSerHisProAspGlyLysValGluLysValTyrLysAsnGlyC

 91 GCCGTGTTATACTGTTTCCCAATGGAACTCGAAAGGAAGTGAGTG
    ysArgValIleLeuPheProAsnGlyThrArgLysGluValSerA

136 CAGATGGGAAGACCATCACTGTCACTTTCTTTAATGGTGACGTGA
    laAspGlyLysThrIleThrValThrPhePheAsnGlyAspValL

181 AGCAGGTCATGCCAGACCAAAGAGTGATCTACTACTATGCAGCTG
    ysGlnValMetProAspGlnArgValIleTyrTyrTyrAlaAlaA

226 CCCAGACCACTCACAGGACATACCCGGAGANCTGGGAAGTCTTAC
    laGlnThrThrHisArgThrTyrProGlu---TrpGluValLeuH

271 ATTTCTCAAGTGGACAAATAGAAAAACATTACCCAGATGGAAGAA
    isPheSerSerGlyGlnIleGluLysHisTyrProAspGlyArgL

316 AAGAAATCACGTTTCCTGACCAGACTGTTAAAAACTTATTTCCTG
    ysGluIleThrPheProAspGlnThrValLysAsnLeuPheProA

361 ATGGACAAGAAGAAAGCATTTCCCAGATGGTACAATTGTCAGAGT
    spGlyGlnGluGluSerIleSerGlnMetValGlnLeuSerGluT

406 ACAACCCCCTCGCAACAAACTCATATAGTTTATTATTGACAAAAA
    yrAsnProLeuAlaThrAsnSerTyrSerLeuLeuLeuThrLysL

451 AAACTAACTACATACTGCCCAGTTCAAGAGACGGGAATACCAGAT
    ysThrAsnTyrIleLeuProSerSerArgAspGlyAsnThrArgT

496 GGCACTGT
    rpHisCys
```

SEQUENCE 14. IP-2


```
  1 GTGGGAAACATGCGGCTGTCGGTCGCTGCAGCATCTCCCATGGCC
      GlyLysHisAlaAlaValGlyArgCysSerIleSerHisGlyA

 46 GCGTATTTCGCCGTATGGGCCTCGGTCCCGAGTCCCGCATCCATC
    rgValPheArgArgMetGlyLeuGlyProGluSerArgIleHisL

 91 TGTTGCGGAACTTGCTCACAGGGCTGGTGCGGCACGAACGCATCG
    euLeuArgAsnLeuLeuThrGlyLeuValArgHisGluArgIleG

136 AGGCACCATGGGCGCGTGTGGACGAAATGAGGGGCTACGCGGAGA
    luAlaProTrpAlaArgValAspGluMetArgGlyTyrAlaGluL

181 AGCTCATCGACTATGGGAAGCTGGGAGACACTAACGAACGAGCCA
    ysLeuIleAspTyrGlyLysLeuGlyAspThrAsnGluArgAlaM

226 TGCGCATGGCTGACTTCTGGCTCACAGAGAAGGATTTGATCCCAA
    etArgMetAlaAspPheTrpLeuThrGluLysAspLeuIleProL

271 AGCTGTTTCAAGTACTGGCCCCTCGGTACAAAGATCAAACTGGGG
    ysLeuPheGlnValLeuAlaProArgTyrLysAspGlnThrGlyG

316 GCTACACAAGAATGCTGCAGATCCCAAATCGGAGTTTGGATCGGG
    lyTyrThrArgMetLeuGlnIleProAsnArgSerLeuAspArgA

361 CCAAGATGGCAGTGATCGAGTATAAAGGGAATTGCCTCCCACCCC
    laLysMetAlaValIleGluTyrLysGlyAsnCysLeuProProL

406 TTGCCTCTTGCCTCCGCAGAGACAAGCCACCTTACACTCCTAAAC
    euAlaSerCysLeuArgArgAspLysProProTyrThrProLysP

451 CAGCTGCTGCAGGGTTTGCGGCAGGACCTTCAGGCAAAGCCAGGA
    roAlaAlaAlaGlyPheAlaAlaGlyProSerGlyLysAlaArg
```

SEQUENCE 15. IP-3

```
   1  AATAGACCAGGCTTGGGCCAGAATGAGAATCTGAGTGCCATTGAG
      AsnArgProGlyLeuGlyGlnAsnGluAsnLeuSerAlaIleGlu

  46  GGGAAAGGCAAGGTGGGGGGACTGAAGACACGCTGCTCTAGCTGC
      GlyLysGlyLysValGlyGlyLeuLysThrArgCysSerSerCys

  91  AACGTTAAGTTTGAGTCTGAAAGTGAACTCCAGAACCACATCCAA
      AsnValLysPheGluSerGluSerGluLeuGlnAsnHisIleGln

 136  ACCATCCACCGAGAGCTCGTGCCAGACAGCAACAGCACACAGTTG
      ThrIleHisArgGluLeuValProAspSerAsnSerThrGlnLeu

 181  AAAACGCCCCAAGTATCACCAATGCCCAGAATCAGTCCCTCCCAG
      LysThrProGlnValSerProMetProArgIleSerProSerGln

 226  TCGGATGAGAAGAAGACCTATCAATGCATCAAGTGTCAGATGGTT
      SerAspGluLysLysThrTyrGlnCysIleLysCysGlnMetVal

 271  TTCTACAATGAATGGGATATTCAGGTTCATGTTGCAAATCACATG
      PheTyrAsnGluTrpAspIleGlnValHisValAlaAsnHisMet
                                    ▼
 316  ATTGATGAAGGACTGAACCATGAATGCAAACTCTGCAGCCAGACC
      IleAspGluGlyLeuAsnHisGluCysLysLeuCysSerGlnThr

 361  TTTGACTCTCCTGCCAAACTCCAGTGCCACCTGATAGAGCACAGC
      PheAspSerProAlaLysLeuGlnCysHisLeuIleGluHisSer

 406  TTCGAAGGGATGGGAGGCACCTTCAAGTGTCCAGTCTGCTTTACA
      PheGluGlyMetGlyGlyThrPheLysCysProValCysPheThr

 451  GTATTTGTTCAAGCAAACAAGTTGCAGCAGCATATTTTCTCTGCC
      ValPheValGlnAlaAsnLysLeuGlnGlnHisIlePheSerAla

 496  CATGGACAAGAAGACAAGATCTATGACTGTACACAATGTCCACAG
      HisGlyGlnGluAspLysIleTyrAspCysThrGlnCysProGln

 541  AAGTTTTTCTTCCAAACAGAGCTGCAGAATCATACAATGACCCAA
      LysPhePhePheGlnThrGluLeuGlnAsnHisThrMetThrGln

 586  CACAGCAGTTAGTGCAAGTACAGTCTCTCAAGGAGAATTGATTTT
      HisSerSer

 631  GTGGCACAAAAAGGGAACATGTTTTACTCTTTGCACGAAACTTTC
 676  ATTGTTAATGTATATTATTCAGAAACATTGTATTGTACCATAAAA
 721  CTTGTATTATCAAACTGTTGGATGTTCATGTGTTTGAACTTTTGC
 766  GCACCGGATAGACCCCTTGTATATAAAGTGTTGCACATGTATTAT
 811  GTCGTCTGATACTAAAATGGTCTTATAAAGACAAGTGGACTTGGG
 856  CCCTATTCAGGCAAGATTAAAAAAAAAAAAAAAGACTATGACCAAA
 901  ATGGCTTAAGATAAAGTATTTTTAAGGAAGAAAGATTAAAAACAA
 946  CTGTTATACATGAGACTATGGTTGGACTTCCTTTTCTTTACACTT
 991  AAGCCTAGAATTTCTCTTTAGGTATATCAGCGCTTAAATCCAAGA
```

```
1036  CTATTTTTATTGCTGAAGATTCTTGCAAACCATGAAGAGATGTT
1081  CTCACAGAACAGAACCCCCACAGCTGGNATAAGGCCCGGTATATA
1126  TATATTTGGTAAAGCCCTTGCAATGTGACAGGGTAGCATCACTAA
1171  TATATGCAATAGGTTGTTATGGAGACTG
```

SEQUENCE 16. IP-4


```
  1 CCTGGCCCAGTACCTCATCACTGCCTGTCTAGAGACGGCCAAGTC
     LeuAlaGlnTyrLeuIleThrAlaCysLeuGluThrAlaLysSe

 46 TACCTCTGGGGGGACGCCCAGCTCGGCCCCCGCAGACCTGCCTGC
    rThrSerGlyGlyThrProSerSerAlaProAlaAspLeuProAl

 91 CCCCTTCTCGCCACCCCTGACGGCCCTGCCACAGCTCCCCCTGGC
    aProPheSerProProLeuThrAlaLeuProGlnLeuProLeuAl

136 CTGCTGGGCACACCCTATGCCATCTCCCTCTCCAACTTCATCGGC
    aCysTrpAlaHisProMetProSerProSerProThrSerSerAl

181 CTCAAGCCCATGCCCTTCTTGGCTTTACCACCTGCTTCCCCAGGG
    aSerSerProCysProSerTrpLeuTyrHisLeuLeuProGlnGl

226 CCGCCGCCGGGCTTGGCGCTAACACTGCCAAGATGGCAGCAGCTA
    yArgArgArgAlaTrpArg

271 ATGGGAGCAAGAAGGCTGAGCGGCAGAAATTCTCCCCCTACTGAG
316 GCCAGCTGAGGTACAGGCAGGGGCAGGCAGGACCACCAGCAGGGG
361 GCTTGCCTCTGCACCCTACCCGCCCAAGGAGACTCCACCCTGGGG
406 TCCCAAACGCCGCTAACGCCCAGACGCATGGATTGCACCCCCTAC
451 CCTGCCTCCATCTATGGGAGTTCTTTCTCTCAGAGTGGGGGCAGT
496 TTCTGGGCCCAGGGGTCTGAGCTGCGGCA
```

SEQUENCE 17. IP-5

```
  1 GTAGAGATGGCCGCGCTTGCACCGCTGCCCCCGCTCCCCGCACAG
    ValGluMetAlaAlaLeuAlaProLeuProProLeuProAlaGln

 46 TTCAAGAGCATACAGCATCATCTGAGGACGGCTCAGGAGCATGAC
    PheLysSerIleGlnHisHisLeuArgThrAlaGlnGluHisAsp

 91 AAGCGAGACCCTGTGGTGGCTTATTACTGTCGTTTATACGCAATG
    LysArgAspProValValAlaTyrTyrCysArgLeuTyrAlaMet

136 CAGACTGGAATGAAGATCGATAGTAAAACTCCTGAATGTCGCAAA
    GlnThrGlyMetLysIleAspSerLysThrProGluCysArgLys

181 TTTTTATCAAAGTTAATGGATCAGTTAGAAGCTCTAAAGAAGCAG
    PheLeuSerLysLeuMetAspGlnLeuGluAlaLeuLysLysGln

226 TTGGGTGATAATGAAGCTATTACTCAAGAAATAGTGGGCTGTGCA
    LeuGlyAspAsnGluAlaIleThrGlnGluIleValGlyCysAla

271 CTTTGGAGAATTATGCTTTGAAAATGTTTTTGTATGCAGACAATG
    LeuTrpArgIleMetLeu

316 AAGATCGTGCTGGACGATTTCACAAAAACATGATCAAGTCCTTCT
361 ATACTGCAAGTCTTTTGATAGATGTCATAACAGTATTTGGAGAAC
406 TCACTGATGAAAATGTGAAACACAGGAAGTATGGCCAGAGGGAAG
451 GCAACATACATCCATAATTGTTTNAAGANTGGGGAGACTCCCAAG
496 CAGGCCCTGNTGGAATTGAAGA
```

SEQUENCE 18. IP-6


```
  1 GAGTATGAGATGAAACGAATGGCAGAGAATGAGCTGAGCCGGTCA
    GluTyrGluMetLysArgMetAlaGluAsnGluLeuSerArgSer

 46 GTAAATGAGTTTCTGTCCAAGCTGCAAGATGACCTCAAGGAGGCA
    ValAsnGluPheLeuSerLysLeuGlnAspAspLeuLysGluAla

 91 ATGAATACTATGATGTGTAGCCGATGCCAAGGAAAGCATAGGAGG
    MetAsnThrMetMetCysSerArgCysGlnGlyLysHisArgArg

136 TTTGAAATGGACCGGGAACCTAAGAGTGCCAGATACTGTGCTGAG
    PheGluMetAspArgGluProLysSerAlaArgTyrCysAlaGlu

181 TGTAATAGGCTGCATCCTGCTGAGGAAGGAGACTTTTGGGCAGAG
    CysAsnArgLeuHisProAlaGluGluGlyAspPheTrpAlaGlu

226 TCAAGCATGTTGGGCCTCAAGATCACCTACTTTGCACTGATGGAT
    SerSerMetLeuGlyLeuLysIleThrTyrPheAlaLeuMetAsp

271 GGAAAGGTGTATGACATCACAGAGTGGGCTGGATGCCAGCGTGTA
    GlyLysValTyrAspIleThrGluTrpAlaGlyCysGlnArgVal

316 GGTATCTCCCCAGATACCCACAGAGTCCCCTATCACATCTCATTT
    GlyIleSerProAspThrHisArgValProTyrHisIleSerPhe

361 GGTTCTCGGATTCCAGGCACCAGAGGGCGGCAGAGAGCCACCCCA
    GlySerArgIleProGlyThrArgGlyArgGlnArgAlaThrPro

406 GATGCCCCTCCTGCTGATCTTCAGGATTTCTTGAGTCGGATCTTT
    AspAlaProProAlaAspLeuGlnAspPheLeuSerArgIlePhe

451 CAAGTACCCCCAGGGCAGATGCCAATGGGAACTTCTTTGCAGCTC
    GlnValProProGlyGlnMetProMetGlyThrSerLeuGlnLeu

496 CTCAGCCTGCCCCTGGAGCCGCTGCAGCCTCTAAGCCCAACAGCA
    LeuSerLeuProLeuGluProLeuGlnProLeuSerProThrAla

541 CAGTACCCAAGGGAGAAGCCAAACCTAAGCGGCGGAAGAAAGTGA
    GlnTyrProArgGluLysProAsnLeuSerGlyGlyArgLys

586 GGAGGCCCTTCCAACGTTGATGCCCCTTCTCTTTCCTCAAATCAA
631 TGTCAGGGAGTCAAAAGGGCTGTAGCACAGGATGGAGTTTGATTT
676 ATCCCTCCTCCCCCAACACCTAGGAACTGAATCTTTTTCTTTTTA
721 TTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAG
```

SEQUENCE 19. IP-7

```
                ↓
1   AAAACGGGTTTTCGCCATGTTGCTCAGGCTGGTCTTGAACTCCTG
                        MetLeuLeuArgLeuValLeuAsnSerTr

46  GGCTCAAGAATCTGTCTGCCTTAGCCAACCAAAGTGCTGGTATTA
    pAlaGlnGluSerValCysLeuSerGlnProLysCysTrpTyrTy

91  CAGATGTGAGCTATTGCACGTGGCTAGGTTTCTCTGCTTTGAAGT
    rArgCysGluLeuLeuHisValAlaArgPheLeuCysPheGluVa

136 TAATCTTTTTCTCCCTTTTCATAAATCTTTGGAAGGAAGTCACTG
    lAsnLeuPheLeuProPheHisLysSerLeuGluGlySerHisCy

181 TGTGAAGCCTACACGTAGGAATGGGAAGTTACCCTGTTCCTCCAA
    sValLysProThrArgArgAsnGlyLysLeuProCysSerSerLy

226 GGTGGTGGAGTATCTACATAATCTTCTGGAATTCTGTTCTGGAGA
    sValValGluTyrLeuHisAsnLeuLeuGluPheCysSerGlyGl

271 GTTGCCTGTTCTCCCCCATTAATTAACTTATATCAGTATGGACNC
    uLeuProValLeuProHis

316 ATGTATATTTACTTTATACCTTGGGTTAAAATCCAATACTAGTTA
361 GATTTGTTGCTCAAATTGTNCCAGCTTTGGCCACTGAAACTCTTT
406 CAGTTGGCAGTGCNGCTGTTAAAAANAAAAAAANAAACCNGTT
```

SEQUENCE 20. IP-8

```
 1 GGCCCCCTCGCTTCTGTGCAAGTATTTCCTGCAGCAGTGGGAACT
    AlaProSerLeuLeuCysLysTyrPheLeuGlnGlnTrpGluLe

46 CACATCCCCTGGCCACGACACCTCGGTGCTGANTGACAGCGTGGA
   uThrSerProGlyHisAspThrSerValLeu---AspSerValGl

91 GATTGGCCTGCAGACCTGCTGCCACATCTTCCTCAACCTCGTGGT
   uIleGlyLeuGlnThrCysCysHisIlePheLeuAsnLeuValVa

136 CACCGCACCGGGGCTGATCAAGCGTGACGCCTGCTTCACATCTCT
    lThrAlaProGlyLeuIleLysArgAspAlaCysPheThrSerLe

181 AATGAACACCCTCATGACGTCGCTACCAGCACTAGTGCAGCAACA
    uMetAsnThrLeuMetThrSerLeuProAlaLeuValGlnGlnGl

226 GGGAAGGCTGCTTCTGGCTGCTAATGTGGCCACCCTGGGGCTCCT
    nGlyArgLeuLeuLeuAlaAlaAsnValAlaThrLeuGlyLeuLe

271 CATGGCCCGGCTCCTTAGCACCTCTCCAGCTCTTCAGGGAACACC
    uMetAlaArgLeuLeuSerThrSerProAlaLeuGlnGlyThrPr

316 AGCATCCCGAGGGTTCTTCGCAGCTGCCATCCTCTTCCTATCACA
    oAlaSerArgGlyPhePheAlaAlaAlaIleLeuPheLeuSerGl

361 GTCCCACGTGGCGCGGGCCACCCCGGGCTCAGACCAGGCAGTGCT
    nSerHisValAlaArgAlaThrProGlySerAspGlnAlaValLe

406 AGCCCTGTCCCCTGAGTATGAGGGCATCTGGGCCGACCTGCAGGA
    uAlaLeuSerProGluTyrGluGlyIleTrpAlaAspLeuGlnGl

451 GCTCTGGTTCCTGGGCATGCAGGCCTTCACCGGCTGTGTGCCTCT
    uLeuTrpPheLeuGlyMetGlnAlaPheThrGlyCysValProLe

496 GCTGCCCTGGCTGGCCCCCGCTGCCCTG
    uLeuProTrpLeuAlaProAlaAla
```

SEQUENCE 21. IP-9

```
   1  GATGAGCATGGCAGTGCCAAGAACGTCAGCTTCAATCCTGCCAAG
      AspGluHisGlySerAlaLysAsnValSerPheAsnProAlaLys

  46  ATCAGTTCCAACTTCAGCAGCATTATTGCAGAGAAATTACGTTGT
      IleSerSerAsnPheSerSerIleIleAlaGluLysLeuArgCys

  91  AATACCCTTCCTGACACTGGTCGCAGGAAGCCCCAAGTGAACCAG
      AsnThrLeuProAspThrGlyArgArgLysProGlnValAsnGln

 136  AAGGATAACTTCTGGCTGGTGACTGCACGATCCCAGAGTGCCATT
      LysAspAsnPheTrpLeuValThrAlaArgSerGlnSerAlaIle

 181  AACACTTGGTTCACTGACTTGGCTGGCACCAAGCCACTCACGCAA
      AsnThrTrpPheThrAspLeuAlaGlyThrLysProLeuThrGln

 226  CTAGCCAAAAAGGTCCCCATTTTCAGTAAGAAGGAAGAGGTGTTT
      LeuAlaLysLysValProIlePheSerLysLysGluGluValPhe

 271  GGGTACTTAGCCAAATACACAGTGCCTGTGATGCGGGCTGCCTGG
      GlyTyrLeuAlaLysTyrThrValProValMetArgAlaAlaTrp

 316  CTCATTAAGATGACCTGTGCCTACTATGCAGCAATCTCTGAGACC
      LeuIleLysMetThrCysAlaTyrTyrAlaAlaIleSerGluThr

 361  AAGGTTAAGAAGAGACATGTTGACCCTTTCATGGAATGGACTCAG
      LysValLysLysArgHisValAspProPheMetGluTrpThrGln

 406  ATCATCACCAAGTACTTATGGGAGCAGTTACAGAAGATGGCTGAA
      IleIleThrLysTyrLeuTrpGluGlnLeuGlnLysMetAlaGlu

 451  TACTACCGGCCAGGGCCTGCAGGAAGTGGGGGCTGTGGTTCCACG
      TyrTyrArgProGlyProAlaGlySerGlyGlyCysGlySerThr

 496  ATAGGGCCCTTGCCCCATGATGTAGAGGTGGCAATCCGGCAGTGG
      IleGlyProLeuProHisAspValGluValAlaIleArgGlnTrp

 541  GATTACACCGAGAAGCTGGCCATGTTCATGTTTCAGGATGGAATG
      AspTyrThrGluLysLeuAlaMetPheMetPheGlnAspGlyMet

 586  CTGGACAGACATGAGTTCCTGACCTGGGTGCTTGAGTGTTTTGAG
      LeuAspArgHisGluPheLeuThrTrpValLeuGluCysPheGlu

 631  AAGATCCGCCCTGGAGAGGATGAATTGCTTAAACTGCTGCTGCCT
      LysIleArgProGlyGluAspGluLeuLeuLysLeuLeuLeuPro

 676  CTGCTTCTCCGATACTCTGGGGAATTTGTTCAGTCTGCATACCTG
      LeuLeuLeuArgTyrSerGlyGluPheValGlnSerAlaTyrLeu

 721  TCCCGCCGGCTTGCCTACTTCTGTACACGGAGACTGGCCCTGCAG
      SerArgArgLeuAlaTyrPheCysThrArgArgLeuAlaLeuGln
```

```
766  CTGGATGGTGTGAGCAGTCACTCATCTCATGTTATATCTGCTCAG
     LeuAspGlyValSerSerHisSerSerHisValIleSerAlaGln

811  TCAACAAGCACGCTACCCACCACCCCTGCTCCTCAGCCCCCAACT
     SerThrSerThrLeuProThrThrProAlaProGlnProProThr

856  AGCAGCACACCCTCGACTCCCTTTAGTGACCTGCTTATGTGCCCT
     SerSerThrProSerThrProPheSerAspLeuLeuMetCysPro

901  CAGCACCGGCCCCTGGTTTTTGGCCTCAGCTGTATCCTACAGACC
     GlnHisArgProLeuValPheGlyLeuSerCysIleLeuGlnThr

946  ATCCTCCTGTGCTGTCCTAGTGCCTTGGTTTGGCACTACTCACTG
     IleLeuLeuCysCysProSerAlaLeuValTrpHisTyrSerLeu

991  ACTGATAGCAGAATTAAGACCGGCTCACCACTTGACCACTTGCCT
     ThrAspSerArgIleLysThrGlySerProLeuAspHisLeuPro

1036 ATTGCCCCGTCCAACCTGCCCATGCCAGAGGGTAACAGTGCCTTC
     IleAlaProSerAsnLeuProMetProGluGlyAsnSerAlaPhe

1081 ACTCAGCAGGTCCGTGCAAAGTTGCGGGAGATCGAGCAGCAGATC
     ThrGlnGlnValArgAlaLysLeuArgGluIleGluGlnGlnIle

1126 AAGGAGCGGGGACAGGCAGTTGAAGTTCGCTGGTCTTTCGATAAA
     LysGluArgGlyGlnAlaValGluValArgTrpSerPheAspLys

1171 TGCCAGGAAGCTACTGCAGGCTTCACCATTGGACGGGTACTTCAT
     CysGlnGluAlaThrAlaGlyPheThrIleGlyArgValLeuHis

1216 ACTTTGGAAGTGCTGGACAGCCATAGTTTTGAACGCTCTGACTTC
     ThrLeuGluValLeuAspSerHisSerPheGluArgSerAspPhe

1261 AGCAACTCTCTTGACTCCCTTTGTAACCGAATCTTTGGATTGGGA
     SerAsnSerLeuAspSerLeuCysAsnArgIlePheGlyLeuGly

1306 CCTAGCAAGGATGGGCATGAGATCTCCTCAGATGATGATGCTGTG
     ProSerLysAspGlyHisGluIleSerSerAspAspAspAlaVal

1351 GTGTCATTGCTATGTGAATGGGCTGTCAGCTGCAAGCGTTCTGGT
     ValSerLeuLeuCysGluTrpAlaValSerCysLysArgSerGly

1396 CGGCATCGTGCTATGGTGGTAGCCAAGCTCCTGGAGAAGAGACAG
     ArgHisArgAlaMetValValAlaLysLeuLeuGluLysArgGln

1441 GCGGAGATTGAGGCTGAGCGTTGTGGAGAATCAGAAGCCGCAGAT
     AlaGluIleGluAlaGluArgCysGlyGluSerGluAlaAlaAsp

1486 GAGAAGGGTTCCATCGCCTCTGGCTCCCTTTCTGCTCCCAGTGCT
     GluLysGlySerIleAlaSerGlySerLeuSerAlaProSerAla

1531 CCCATTTTCCAGGATGTCCTCCTGCAGTTTCTGGATACACAGGCT
     ProIlePheGlnAspValLeuLeuGlnPheLeuAspThrGlnAla
```

```
1576  CCCATGCTGACGGACCCTCGAAGTGAGAGTGAGCGGGTGGAATTC
      ProMetLeuThrAspProArgSerGluSerGluArgValGluPhe

1621  TTTAACTTAGTACTGCTGTTCTGTGAACTGATTCGACATGATGTT
      PheAsnLeuValLeuLeuPheCysGluLeuIleArgHisAspVal

1666  TTCTCCCACAACATGTATACTTGCACTCTCATCTCCCGAGGGGAC
      PheSerHisAsnMetTyrThrCysThrLeuIleSerArgGlyAsp

1711  CTTGCCTTTGGAGCCCCTGGTCCCCGGCCTCCCTCTCCCTTTGAT
      LeuAlaPheGlyAlaProGlyProArgProProSerProPheAsp

1756  GATCCTGCCGATGACCCAGAGCACAAGGAGGCTGAAGGCAGCAGC
      AspProAlaAspAspProGluHisLysGluAlaGluGlySerSer

1801  AGCAGCAAGCTGGAAGATCCAGGGCTCTCAGAATCTATGGACATT
      SerSerLysLeuGluAspProGlyLeuSerGluSerMetAspIle

1846  GACCCTAGTTCCAGTGTTCTCTTTGAGGACATGGAGAAGCCTGAT
      AspProSerSerSerValLeuPheGluAspMetGluLysProAsp

1891  TTCTCATTGTTCTCCCCTACTATGCCCTGTGAGGGGAAGGGCAGT
      PheSerLeuPheSerProThrMetProCysGluGlyLysGlySer

1936  CCATCCCCTGAGAAGCCAGATGTCGAGAAGGAGGTGAAGCCCCCA
      ProSerProGluLysProAspValGluLysGluValLysProPro

1981  CCCAAGGAGAAGATTGAAGGGACCCTTGGGGTTCTTTACGACCAG
      ProLysGluLysIleGluGlyThrLeuGlyValLeuTyrAspGln

2026  CCACGACACGTGCAGTACGCCACCCATTTTCCCATCCCCCAGGAG
      ProArgHisValGlnTyrAlaThrHisPheProIleProGlnGlu

2071  GAGTCATGCAGCCATGAGTGCAACCAGCGGTTGGTCGTACTGTTT
      GluSerCysSerHisGluCysAsnGlnArgLeuValValLeuPhe

2116  GGGGTGGGAAAGCAGCGAGATGATGCCCGCCATGCCATCAAGAAA
      GlyValGlyLysGlnArgAspAspAlaArgHisAlaIleLysLys

2161  ATCACCAAGGATATCTTGAAGGTTCTGAACCGCAAAGGGACAGCA
      IleThrLysAspIleLeuLysValLeuAsnArgLysGlyThrAla

2206  GAAACTGACCAGCTTGCTCCTATTGTGCCTCTGAATCCTGGAGAC
      GluThrAspGlnLeuAlaProIleValProLeuAsnProGlyAsp

2251  CTGACATTCTTAGGTGGGGAGGATGGGCAGAAGCGGCGACGCAAC
      LeuThrPheLeuGlyGlyGluAspGlyGlnLysArgArgArgAsn

2296  CGGCCTGAAGCCTTCCCCACTGCTGAAGATATCTTTGCTAAGTTC
      ArgProGluAlaPheProThrAlaGluAspIlePheAlaLysPhe

2341  CAGCACCTTTCACATTATGACCAACACCAGGTCACGGCTCAGGTC
      GlnHisLeuSerHisTyrAspGlnHisGlnValThrAlaGlnVal
```

```
2386  TCCCGGAATGTTCTGGAGCAGATCACGAGCTTTGCCCTTGGCATG
      SerArgAsnValLeuGluGlnIleThrSerPheAlaLeuGlyMet

2431  TCATACCACTTGCCTCTGGTGCAGCATGTGCAGTTCATCTTCGAC
      SerTyrHisLeuProLeuValGlnHisValGlnPheIlePheAsp

2476  CTCATGGAATATTCACTCAGCATCAGTGGCCTCATCGACTTTGCC
      LeuMetGluTyrSerLeuSerIleSerGlyLeuIleAspPheAla

2521  ATTCAGCTGCTGAATGAACTGAGTGTAGTTGAGGCTGAGCTGCTT
      IleGlnLeuLeuAsnGluLeuSerValValGluAlaGluLeuLeu

2566  CTCAAATCCTCGGATCTGGTGGGCAGCTACACTACTAGCCTGTGC
      LeuLysSerSerAspLeuValGlySerTyrThrThrSerLeuCys

2611  CTGTGCATCGTGGCTGTCCTGCGGCACTATCATGCCTGCCTCATC
      LeuCysIleValAlaValLeuArgHisTyrHisAlaCysLeuIle

2656  CTCAACCAGGACCAGATGGCACAGGTCTTTGAGGGGCTGTGTGGC
      LeuAsnGlnAspGlnMetAlaGlnValPheGluGlyLeuCysGly

2701  GTCGTGAAGCATGGGATGAACCGGTCCGATGGCTCCTCTGCAGAG
      ValValLysHisGlyMetAsnArgSerAspGlySerSerAlaGlu

2746  CGCTGTATCCTTGCTTATCTCTATGATCTGTACACCTCCTGTAGC
      ArgCysIleLeuAlaTyrLeuTyrAspLeuTyrThrSerCysSer

2791  CATTTAAAGAACAAATTTGGGGAGCTCTTCAGCGACTTTTGCTCA
      HisLeuLysAsnLysPheGlyGluLeuPheSerAspPheCysSer

2836  AAGGTGAAGAACACCATCTACTGCAACGTGGAGCCATCGGAATCA
      LysValLysAsnThrIleTyrCysAsnValGluProSerGluSer

2881  AATATGCGCTGGGCACCTGAGTTCATGATCGACACTCTAGAGAAC
      AsnMetArgTrpAlaProGluPheMetIleAspThrLeuGluAsn

2926  CCTGCAGCTCACACCTTCACCTACACGGGGCTAGGCAAGAGTCTT
      ProAlaAlaHisThrPheThrTyrThrGlyLeuGlyLysSerLeu

2971  AGTGAGAACCCTGCTAACCGCTACAGCTTTGTCTGCAATGCCCTT
      SerGluAsnProAlaAsnArgTyrSerPheValCysAsnAlaLeu

3016  ATGCACGTCTGTGTGGGGCACCATGATCCCGATAGGGTGAATGAC
      MetHisValCysValGlyHisHisAspProAspArgValAsnAsp

3061  ATCGCAATCCTGTGTGCAGAGCTGACCGGCTATTGCAAGTCACTG
      IleAlaIleLeuCysAlaGluLeuThrGlyTyrCysLysSerLeu

3106  AGTGCAGAATGGCTAGGAGTGCTTAAGGCCTTGTGCTGCTCCTCT
      SerAlaGluTrpLeuGlyValLeuLysAlaLeuCysCysSerSer

3151  AACAATGGCACTTGTGGTTTCAACGATCTCCTCTGCAATGTTGAT
      AsnAsnGlyThrCysGlyPheAsnAspLeuLeuCysAsnValAsp
```

```
3196  GTCAGTGACCTATCTTTTCATGACTCGCTGGCTACTTTTGTTGCC
      ValSerAspLeuSerPheHisAspSerLeuAlaThrPheValAla

3241  ATCCTCATCGCTCGGCAGTGTTTGCTCCTGGAAGATCTGATTCGC
      IleLeuIleAlaArgGlnCysLeuLeuLeuGluAspLeuIleArg

3286  TGTGCTGCCATCCCTTCACTCCTTAATGCTGCTTGTAGTGAACAG
      CysAlaAlaIleProSerLeuLeuAsnAlaAlaCysSerGluGln

3331  GACTCTGAGCCAGGGGCCCGGCTTACCTGCCGCATCCTCCTTCAC
      AspSerGluProGlyAlaArgLeuThrCysArgIleLeuLeuHis

3376  CTTTTCAAGACACCGCAGCTCAATCCTTGCCAGTCTGATGGAAAC
      LeuPheLysThrProGlnLeuAsnProCysGlnSerAspGlyAsn

3421  AAGCCTACAGTAGGAATCCGCTCCTCCTGCGACCGCCACCTGCTG
      LysProThrValGlyIleArgSerSerCysAspArgHisLeuLeu

3466  GCTGCCTCCCAGAACCGCATCGTGGATGGAGCCGTGTTTGCTGTT
      AlaAlaSerGlnAsnArgIleValAspGlyAlaValPheAlaVal

3511  CTCAAGGCTGTGTTTGTACTTGGGGATGCGGAACTGAAAGGTTCA
      LeuLysAlaValPheValLeuGlyAspAlaGluLeuLysGlySer

3556  GGCTTCACTGTGACAGGAGGAACAGAAGAACTTCCAGAGGAGGAG
      GlyPheThrValThrGlyGlyThrGluGluLeuProGluGluGlu

3601  GGAGGAGGTGGCAGTGGTGGTCGGAGGCAGGGTGGCCGCAACATC
      GlyGlyGlyGlySerGlyGlyArgArgGlnGlyGlyArgAsnIle

3646  TCTGTGGAGACAGCCAGTCTGGATGTCTATGCCAAGTACGTGCTG
      SerValGluThrAlaSerLeuAspValTyrAlaLysTyrValLeu

3691  CGCAGCATCTGCCAACAGGAATGGGTAGGAGAACGTTGCCTTAAG
      ArgSerIleCysGlnGlnGluTrpValGlyGluArgCysLeuLys

3736  TCTCTGTGTGAGGACAGCAATGACCTGCAAGACCCAGTGTTGAGT
      SerLeuCysGluAspSerAsnAspLeuGlnAspProValLeuSer

3781  AGTGCCCAGGCGCAGCGCCTCATGCAGCTCATTTGCTATCCACAT
      SerAlaGlnAlaGlnArgLeuMetGlnLeuIleCysTyrProHis

3826  CGACTGCTGGACAATGAGGATGGGGAAAACCCCCAGCGGCAGCGC
      ArgLeuLeuAspAsnGluAspGlyGluAsnProGlnArgGlnArg

3871  ATAAAGCGCATTCTCCAGAACTTGGACCAGTGGACCATGCGCCAG
      IleLysArgIleLeuGlnAsnLeuAspGlnTrpThrMetArgGln

3916  TCTTCCTTGGAGCTGCAGCTCATGATCAAGCAGACCCCTAACAAT
      SerSerLeuGluLeuGlnLeuMetIleLysGlnThrProAsnAsn

3961  GAGATGAACTCCCTCTTGGAGAACATCGCCAAGGCCACAATCGAG
      GluMetAsnSerLeuLeuGluAsnIleAlaLysAlaThrIleGlu
```

```
4006  GTTTTCCAACGGTCAGCAGAGACAGGGTCATCTTCTGGAAGTACT
      ValPheGlnArgSerAlaGluThrGlySerSerSerGlySerThr

4051  GCAAGCAACATGCCCAGCAGCAGCAAGACCAAGCCTGTGCTCAGC
      AlaSerAsnMetProSerSerSerLysThrLysProValLeuSer

4096  TCTCTAGAGCGCTCTGGTGTATGGCTGGTGGCCCCCCTCATTGCT
      SerLeuGluArgSerGlyValTrpLeuValAlaProLeuIleAla

4141  AAACTGCCCACCTCAGTCCAGGGACATGTGTTAAAGGCTGCTGGG
      LysLeuProThrSerValGlnGlyHisValLeuLysAlaAlaGly

4186  GAAGAATTGGAGAAGGGTCAGCACCTGGGTTCCTCTTCACGCAAA
      GluGluLeuGluLysGlyGlnHisLeuGlySerSerSerArgLys

4231  GAACGTGATCGACAAAAGCAGAAGAGCATGTCCCTATTGAGCCAG
      GluArgAspArgGlnLysGlnLysSerMetSerLeuLeuSerGln

4276  CAGCCCTTCTTATCGCTGGTGCTAACATGTCTGAAAGGGCAGGAT
      GlnProPheLeuSerLeuValLeuThrCysLeuLysGlyGlnAsp

4321  GAACAACGCGAGGGACTCCTTACCTCCCTCTACAGCCAGGTGCAC
      GluGlnArgGluGlyLeuLeuThrSerLeuTyrSerGlnValHis

4366  CAGATTGTGAATAATTGGCGAGATGACCAGTACTTAGATGATTGC
      GlnIleValAsnAsnTrpArgAspAspGlnTyrLeuAspAspCys

4411  AAACCAAAGCAGCTTATGCATGAGGCACTCAAACTGCGGCTCAAC
      LysProLysGlnLeuMetHisGluAlaLeuLysLeuArgLeuAsn

4456  CTGGTGGGGGGCATGTTTGACACGGTGCAGCGCAGCACCCAGCAG
      LeuValGlyGlyMetPheAspThrValGlnArgSerThrGlnGln

4501  ACCACGGAGTGGGCCATGCTCCTCCTGGAGATCATCATCAGCGGC
      ThrThrGluTrpAlaMetLeuLeuGluIleIleIleSerGly

4546  ACTGTCGACATGCAGTCCAACAATGAGCTCTTCACTACTGTGTTG
      ThrValAspMetGlnSerAsnAsnGluLeuPheThrThrValLeu

4591  GACATGCTGAGCGTGCTCATCAATGGGACATTGGCTGCAGACATG
      AspMetLeuSerValLeuIleAsnGlyThrLeuAlaAlaAspMet

4636  TCTAGCATCTCGCAAGGTAGCATGGAGGAAAACAAGCGTGCATAC
      SerSerIleSerGlnGlySerMetGluGluAsnLysArgAlaTyr

4681  ATGAACCTGGCGAAGAAGTTGCAGAAGGAGTTGGGGGAGCGCCAG
      MetAsnLeuAlaLysLysLeuGlnLysGluLeuGlyGluArgGln

4726  TCAGACAGTCTGGAAAAGGTTCGCCAGCTGCTGCCACTGCCCAAG
      SerAspSerLeuGluLysValArgGlnLeuLeuProLeuProLys

4771  CAGACCCGAGATGTCATCACGTGTGAGCCACAGGGCTCCCTTATC
      GlnThrArgAspValIleThrCysGluProGlnGlySerLeuIle
```

```
4816  GATACCAAGGGCAACAAGATTGCTGGCTTCGATTCCATCTTCAAG
      AspThrLysGlyAsnLysIleAlaGlyPheAspSerIlePheLys

4861  AAGGAGGGTCTACAGGTTTCCACCAAACAGAAGATCTCGCCCTGG
      LysGluGlyLeuGlnValSerThrLysGlnLysIleSerProTrp

4906  GATCTTTTTGAGGGGTTGAAGCCGTCAGCACCACTCTCTTGGGGC
      AspLeuPheGluGlyLeuLysProSerAlaProLeuSerTrpGly

4951  TGGTTTGGAACAGTCCGAGTGGACCGGCGAGTGGCTCGAGGAGAG
      TrpPheGlyThrValArgValAspArgArgValAlaArgGlyGlu

4996  GAGCAGCAGCGGTTGCTGCTCTACCACACACACCTGAGGCCCCGG
      GluGlnGlnArgLeuLeuLeuTyrHisThrHisLeuArgProArg

5041  CCCCGCGCCTATTACCTGGAGCCACTGCCACTGCCCCCAGAAGAT
      ProArgAlaTyrTyrLeuGluProLeuProLeuProProGluAsp

5086  GAGGAGCCGCCTGCTCCCTACCCTGCTAGAGCCTGAGAAAAAGGCT
      GluGlu**Pro**ProAlaProThrLeuLeuGluProGluLysLysAla

5131  CCAGAGCCCCCCAAAACTGACAAACCGGGGGCTGCTCCACCCAGT
      ProGluProProLysThrAspLysProGlyAlaAlaProProSer

5176  ACTGAGGAACGCAAGAAGAAGTCCACCAAGGGCAAGAAACGCAGC
      ThrGluGluArgLysLysLysSerThrLysGlyLysLysArgSer

5221  CAGCCAGCTACCAAGACAGAGGACTATGGAATGGGCCCGGGTCGG
      GlnProAlaThrLysThrGluAspTyrGlyMetGlyProGlyArg

5266  AGCGGCCCTTATGGTGTGACAGTGCCTCCGGACCTCCTGCACCAC
      SerGlyProTyrGlyValThrValProProAspLeuLeuHisHis

5311  CCAAACCCTGGTTCTATAACACACCTTAACTACAGGCAAGGCTCC
      ProAsnProGlySerIleThrHisLeuAsnTyrArgGlnGlySer

5356  ATAGGCCTGTACACCCAGAACCAGCCACTACCTGCAGGTGGCCCT
      IleGlyLeuTyrThrGlnAsnGlnProLeuProAlaGlyGlyPro

5401  CGTGTGGACCCATACCGTCCTGTGCGCTTACCAATGCAGAAGCTG
      ArgValAspProTyrArgProValArgLeuProMetGlnLysLeu

5446  CCCACCCGACCAACTTACCCTGGAGTGCTGCCCACAACCATGACT
      ProThrArgProThrTyrProGlyValLeuProThrThrMetThr

5491  GGCGTCATGGGTTTAGAACCCTCCTCTTATAAGACCTCTGTGTAC
      GlyValMetGlyLeuGluProSerSerTyrLysThrSerValTyr

5536  CGGCAGCAGCAACCTGCGGTGCCCCAAGGACAGCGCCTTCGCCAA
      ArgGlnGlnGlnProAlaValProGlnGlyGlnArgLeuArgGln

5581  CAGCTCCAGCAGAGTCAGGGCATGTTGGGACAGTCATCTGTCCAT
      GlnLeuGlnGlnSerGlnGlyMetLeuGlyGlnSerSerValHis
```

```
5626  CAGATGACTCCCAGCTCTTCCTACGGTTTGCAGACTTCCCAGGGC
      GlnMetThrProSerSerSerTyrGlyLeuGlnThrSerGlnGly

5671  TATACTCCTTATGTTTCTCATGTGGGATTGCAGCAACACACAGGC
      TyrThrProTyrValSerHisValGlyLeuGlnGlnHisThrGly

5716  CCTGCAGGTACCATGGTGCCCCCCAGCTACTCCAGCCAGCCTTAC
      ProAlaGlyThrMetValProProSerTyrSerSerGlnProTyr

5761  CAGAGCACCCACCCTTCTACCAATCCTACTCTTGTAGATCCTACC
      GlnSerThrHisProSerThrAsnProThrLeuValAspProThr

5806  CGCCACCTGCAACAGCGGCCCAGTGGCTATGTGCACCAGCAGGCC
      ArgHisLeuGlnGlnArgProSerGlyTyrValHisGlnGlnAla

5851  CCCACCTATGGACATGGACTGACCTCCACTCAAAGGTTTTCACAC
      ProThrTyrGlyHisGlyLeuThrSerThrGlnArgPheSerHis

5896  CAGACACTGCAGCAGACACCCATGATAAGTACCATGACTCCAATG
      GlnThrLeuGlnGlnThrProMetIleSerThrMetThrProMet

5941  AGTGCCCAGGGCGTCCAGGCAGGCGTCCGTTCAACAGCCATCCTA
      SerAlaGlnGlyValGlnAlaGlyValArgSerThrAlaIleLeu

5986  CCTGAGCAGCAGCAGCAGCAGCAACAGCAGCAACAGCAACAGCAG
      ProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln

6031  CAGCAGCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGTACCAC
      GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnTyrHis

6076  ATCCGGCAGCAGCAGCAGCAGATCCTGCGGCAGCAGCAGCAA
      IleArgGlnGlnGlnGlnGlnGlnIleLeuArgGlnGlnGlnGln

6121  CAGCAACAGCAGCAGCAGCAGCAGCAACAGCAACAGCAGCAG
      GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln

6166  CAGCAACAGCAACAACAGCAACACCAGCAGCAACAGCAGCAACAG
      GlnGlnGlnGlnGlnGlnGlnHisGlnGlnGlnGlnGlnGlnGln

6211  GCGGCTCCTCCCCAACCCCAGCCCCAGTCCCAGCCCCAGTTCCAG
      AlaAlaProProGlnProGlnProGlnSerGlnProGlnPheGln

6256  CGCCAGGGGCTTCAGCAGACCCAGCAGCAGCAACAGACAGCAGCT
      ArgGlnGlyLeuGlnGlnThrGlnGlnGlnGlnGlnThrAlaAla

6301  TTGGTCCGGCAACTTCAACAACAGCTCTCTAATACCCAGCCACAG
      LeuValArgGlnLeuGlnGlnGlnLeuSerAsnThrGlnProGln

6346  CCCAGTACCAACATATTTGGACGCTACTGAGCCACCTGGAGGAAC
      ProSerThrAsnIlePheGlyArgTyr
6391  TGCTTGTGCACTGGATGTGGCCCCACCCTTTCCTCTTAATTCCCA
6436  ATCCCATTCCTGGGCTAGCACCAGTAGTGGTTGGGGCCCTCCCCT
6481  CAGGCTCCATTTTTAATAAGTTTTTAGTATTTTTGTTAATGTGAG
6526  GCATTGAGCTGTTGGGTTTTGTATATTATTTATATAGAGACCCCA
6571  GAGCTGTTGCACCCAATACACAGAGCTTCTTTGC
```

SEQUENCE 22. IP-10

```
  1 GTTTAAAGCTGAGAACGCATCTTTAGCTAAACTTCGCATTGAACG
    PheLysAlaGluAsnAlaSerLeuAlaLysLeuArgIleGluAr

 46 AGAAAGTGCCTTGGAAAAACTCAGGAAAGAAATTGCAGACTTCGA
    gGluSerAlaLeuGluLysLeuArgLysGluIleAlaAspPheGl

 91 ACAACAGAAAGCAAAAGAATTAGCTCGAATAGAAGAGTTTAAAAA
    uGlnGlnLysAlaLysGluLeuAlaArgIleGluGluPheLysLy

136 GGAGGAGATGAGGAAGCTACAAAAGGAACGTAAAGTTTTTGAAAA
    sGluGluMetArgLysLeuGlnLysGluArgLysValPheGluLy

181 GTATACTACAGCTGCAAGAACTTTTCCAGATAAAAAGGAACGTGA
    sTyrThrThrAlaAlaArgThrPheProAspLysLysGluArgGl

226 AGAAATACAGACTTTAAAACAGCAAATAGCAGATTTACGGGAAGA
    uGluIleGlnThrLeuLysGlnGlnIleAlaAspLeuArgGluAs

271 TTTGAAAAGAAAGGAGACCAAATGGTCAAGTACACACAGCCGTCT
    pLeuLysArgLysGluThrLysTrpSerSerThrHisSerArgLe

316 CAGAAGCCAGATACAAATGTTAGTCAGAGAGAACACAGACCTCCG
    uArgSerGlnIleGlnMetLeuValArgGluAsnThrAspLeuAr

361 GGAAGAAATAAAAGTGATGGAAAGATTCCGACTGGATGTCTGGAA
    gGluGluIleLysValMetGluArgPheArgLeuAspValTrpLy

406 GAGAGCAGAAGCCATAGAGAGCAGCCTCGAGGTGGAGGAGGAGGG
    sArgAlaGluAlaIleGluSerSerLeuGluValGluGluGluGl

451 C
    y
```

SEQUENCE 23. IP-11

```
  1 AATTCGCGGNGGCTCGACGCCGCGCCACCGGCCCGGCAGGTGCTG
    PheAla---AlaArgArgArgAlaThrGlyProAlaGlyAlaV

 46 TCCTTATTCCCAGCCCAGTCAAGAGCTACCGGGGCTGGCTAGTCA
    alLeuIleProSerProValLysSerTyrArgGlyTrpLeuValM

 91 TGGGGGAGCCCAGTAGAGAGGAGTATAAAATCCAGTCCTTTGATG
    etGlyGluProSerArgGluGluTyrLysIleGlnSerPheAspA

136 CAGAGACCCAGCAGCTGCTGAAGACAGCACTCAAAGATCCGGGTG
    laGluThrGlnGlnLeuLeuLysThrAlaLeuLysAspProGlyA

181 CTGTGGACTTGGAGAAAGTGGCCAATGTGATTGTGGACCATTCTC
    laValAspLeuGluLysValAlaAsnValIleValAspHisSerL

226 TGCAGGACTGTGTGTTCAGCAAGGAAGCAGGACGCATGTGCTACG
    euGlnAspCysValPheSerLysGluAlaGlyArgMetCysTyrA

271 CCATCATTCAGGCAGAGAGTAAACAAGCAGGCCAGAGTGTCTTCC
    laIleIleGlnAlaGluSerLysGlnAlaGlyGlnSerValPheA

316 GACGTGGACTCCTCAACCGGCTGCAGCAGGAGTACCAGGCTCGGG
    rgArgGlyLeuLeuAsnArgLeuGlnGlnGluTyrGlnAlaArgG

361 AGCAGCTGCGAGCACGCTCCCTGCAGGGCTGGGTCTGCTATGTCA
    luGlnLeuArgAlaArgSerLeuGlnGlyTrpValCysTyrValT

406 CCTTTATCTGCAACATCTTTGACTACCTGAGGGTGAACAACATGC
    hrPheIleCysAsnIlePheAspTyrLeuArgValAsnAsnMetP

451 CCATGATGGCCCTGGTGAACCCTGTCTATGACTGCCTCTTCCGGC
    roMetMetAlaLeuValAsnProValTyrAspCysLeuPheArgL

496 TT
    eu
```

SEQUENCE 24. IP-12

```
                ↓
  1 CTTCTTCTCTCTTTGCTTCAGACTGGCCCGGGAGCAAGCGCGAGT
     PhePheSerLeuCysPheArgLeuAlaArgGluGlnAlaArgVa

 46 GTGCGAACTGCAGAGTGGGAACCAGCAGCTGGAGGAGCAGCGGGT
    lCysGluLeuGlnSerGlyAsnGlnGlnLeuGluGluGlnArgVa

 91 GGAGCTGGTGGAAAGACTGCAGGCCATGCTGCAGGCCCACTGGGA
    lGluLeuValGluArgLeuGlnAlaMetLeuGlnAlaHisTrpAs

136 TGAGGTCAACCAGCTGCTCAGCACCACTCTCCCGCCGCCCAACCC
    pGluValAsnGlnLeuLeuSerThrThrLeuProProProAsnPr

181 TCAAGCTCCTCCTGATGGACCCTCCAGCCCCGGGCCTCAGGAGAC
    oGlnAlaProProAspGlyProSerSerProGlyProGlnGluTh

226 GGAGAAGGAGGAGAGGAGGGTTTGGACTATCCCTCCCATGGCCGT
    rGluLysGluGluArgArgValTrpThrIleProProMetAlaVa

271 GGCCTCAACCCTGTATTCCAACAGAGCCGGGAAGCAAGGCACGAG
    lAlaSerThrLeuTyrSerAsnArgAlaGlyLysGlnGlyThrSe

316 TTACCTGGACCGCCTCCTGTTCTTTTCAATTTCTTCTCAGATCTT
    rTyrLeuAspArgLeuLeuPhePheSerIleSerSerGlnIleLe

361 AGCCTCCTGTTGGGCCCCTCTTTTCATAATCAACATTTTTTC
    uAlaSerCysTrpAlaProLeuPheIleIleAsnIlePhe
```

SEQUENCE 25. IP-13

```
  1 GCGAGCGTGAGTCAGCGCTGAAGGCCCAGAGCGCGCTGCACGAGC
    GluArgGluSerAlaLeuLysAlaGlnSerAlaLeuHisGluG

 46 AGAAGACTCTGCCCGGGATGAACCGGCCGATCCAGGTGAAGCTGC
    lnLysThrLeuProGlyMetAsnArgProIleGlnValLysLeuA

 91 GAACACGAGACCGAGGAGATAGAAAACTCTTCGTGGGCATGCTCA
    rgThrArgAspArgGlyAspArgLysLeuPheValGlyMetLeuA

136 ACAAGCAAAAGTCCGAGGACGACGTGCGCCGCCTTTTCGAGGCCT
    snLysGlnLysSerGluAspAspValArgArgLeuPheGluAlaP

181 TTGGGAACATCGAGGAGTGCACCATCCTGCGCGGGCCCGACGGCA
    heGlyAsnIleGluGluCysThrIleLeuArgGlyProAspGlyA

226 ACAGCAAGGGGTGCGCCTTTGTGAAGTACTCCTCCCACGCCGAGG
    snSerLysGlyCysAlaPheValLysTyrSerSerHisAlaGluG

271 GGCAGGCCGCCATTCAACGCGCTACACGGCAGCCAGACCATGCCG
    lyGlnAlaAlaIleGlnArgAlaThrArgGlnProAspHisAlaG

316 GGAGCCTCGTCCAGTCTGGTGGTCAAGTTCGCCGACACCGACAAG
    lySerLeuValGlnSerGlyGlyGlnValArgArgHisArgGlnG

361 GAGCGCACGATGCGGCGAATGCAGCAGATGGTTGGCCAGATGGGC
    lyAlaHisAspAlaAlaAsnAlaAlaAspGlyTrpProAspGlyH

406 ATGTTCAACCCCATGGCCATCCCTTTCGGGGCCTACGGGCCTACG
    isValGlnProHisGlyHisProPheArgGlyLeuArgAlaTyrA

451 CTCAGGCAATGCAGCAGCAAGCGGCCCTGATGGCATCAGTCGCGC
    laGlnAlaMetGlnGlnGlnAlaAlaLeuMetAlaSerValAlaG

496 AGGGGGCCTACCTGAACCCCATGGCTGCCTTCGCTGTCGCCCAGA
    lnGlyAlaTyrLeuAsnProMetAlaAlaPheAlaValAlaGlnM

541 TGCAGCAGATGGCG
    etGlnGlnMetAla
```

BIBLIOGRAPHY

Aberanthy CR, Rasmussen SA, Stalker HJ, Zori R, Driscoll DJ, Williams CA, Kousseff BG, Wallace MR. (1997) NF1 mutation analysis using a combined heteroduplex/SSCP approach. *Hum Mutation* 9: 548-554.

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377 (6547 Suppl): 3-174.

Adams MD, Soares MB, Kerlavage AR, Fields C, Venter JC. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genetics* 4 (4): 373-380.

Aitken A, Collinge DB, van Heusden BP, Isobe T, Roseboom PH, Rosenfeld G, Soll J. (1992) 14-3-3 proteins: a highly conserved, widespread family of eukaryotic proteins. *Trends Biochem Sci* 17(12): 498-501.

Aitken A, Jones D, Soneji Y, Howell S. (1995) 14-3-3 proteins: biological function and domain structure. *Biochemical Society Transactions* 23(3): 605-11.

Alberts B, Bray D, Lewis J, Raff, M, Roberts K, Watson JD. (1994) Molecular biology of the cell. Garland Publishing, Inc., New York, NY.

Allende JE, Allende CC. (1995) Protein kinase CK2: an enzyme with multiple substrates and a puzzling regulation. *FASEB J* 9: 313-23.

Altschul SF, Gish W, Miller H, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J Mol Biol*, 215: 403-410.

Apitz-Castro R, Cruz MR, Ledezma E, Merino F, Ramirez-Duque P, Dangelmeier C, Holmsen H. (1985) The storage pool deficiency in platelets from humans with the Chediak-Higashi syndrome: study of six patients. *Br J Hematol* 59(3): 471-83.

Arden SD, Roep BO, Neophytou PI, Usac EF, Duinkerken G, de Vries RRP, Hutton JC. (1996) Imogen 38: a novel 38-kD islet mitochondrial autoantigen recognized by T cells from a newly diagnosed type 1 diabetic patient. *J Clin Invest* 97: 551-61.

Asao H, Sasaki Y, Arita T, Tanaka N, Endo K, Kasai H, Takeshita T, Endo Y, Fujita T, Sugamura K. (1997) Hrs is associated with STAM, a signal transducing adaptor molecule. *J Biol Chem* 272: 32785-91.

Baetz K, Isaaz S, Griffiths GM. (1995) Loss of cytotoxic T lymphocyte function in Chediak-Higashi syndrome arises from a secretory defect that prevents lytic granule exocytosis. *J Immunol* 154: 6122-31.

Bairoch A. (1992) PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Research* 20 Suppl:2013-2018.

Ballard R, Tien RD, Nohria V, Juel V. (1994) The Chediak-Higashi syndrome: CT and MR findings. *Pediatr Radiol* 24(4), 266-7.

Barbosa MDFS, Barrat FJ, Tchernev VT, Nguyen QA, Mishra VS, Colman SD, Pastural E, Dufourcq-Lagelouse R, Fischer A, Holcombe RF, Wallace MR, Brandt SJ, de Saint Basile G, Kingsmore SF. (1997). Identification of mutations in two major mRNA isoforms of the Chediak-Higashi syndrome gene in human and mouse. *Hum Mol Genet* 6,1091-1098.

Barbosa MDFS, Nguyen QA, Tchernev VT, Ashley JA, Detter CD, Blaydes SM, Brandt SJ, Chotai D, Hodgman C, Solari RCE, Lovett M, Kingsmore, SF. (1996) Identification of the homologous beige and Chediak-Higashi syndrome genes. *Nature* 382:262-5.

Barrat FJ, Auloge L, Pastural E, Dufourcq-Lagelouse R, Vilmer E, Cant AJ, Weissenbach J, Le Paslier D, FischerA, de Saint Basile G. (1996) Genetic and physical mapping of the Chediak-Higashi syndrome on Chromosome 1q42-43. *Am J Hum Genet* 59:625-32.

Bean AJ, Seifert R, Chen YA, Sacks R, Scheller RH. (1997) Hrs-2 is an ATPase implicated in calcium-regulated sectretion. *Nature*385: 826-29.

Beguez-Cesar A. (1943) Neutropenia cronica maligna familiar con granulaciones atipicas de los leucocitos. *Bol Soc Cubana Pediat.* 15, 900-922.

Beidler JL, Hilliard PR, Rill RL. (1982) Ultrasensitive staining of nucleic acids with silver. *Anal Biochem* 126: 374-80.

Bellinati-Pires R, Salgado MM, Joazerio PP, Carneiro-Sampaio MM. (1992) Delayed phagocytosis and bacterial killing in Chediak-Higashi syndrome neutrophils detected by a fluorochrome assay. *Mem Inst Oswaldo Cruz* 87(4):575-81.

Belmont LD, Mitchison TJ. (1996) Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules. *Cell* 84:623-31.

Benfenati F, Valtorta F, Chieregatti E, Greengard P. (1992) Interaction of free and synaptic vesicle-bound synapsin I with F-actin. *Neuron* 8: 377-86.

Bennet MK, Miller KG, Scheller RH. (1993) Casein kinase II phosphorylates the synaptic vesicle protein p65. *J Neuroscience* 13: 1701-07.

Bishop DT. (1985) The information content of phase-known matings for ordering genetic loci. *Genet Epidemiol*, 2, 349-361.

Blume RS, Bennett JM, Yankee RA, Wolff SM. (1968) Defective granulocyte regulation in the Chediak-Higashi syndrome. *N Engl J Med,* 279(19): 1009-15.

Blume RS, Wolff SM. (1972). The Chediak-Higashi syndrome: studies in four patients and a review of the literature. *Medicine Baltimore* 51(4), 247-80.

Brahmi Z. (1983) Nature of natural killer cell hyporesponsiveness in the Chediak-Higashi syndrome. *Hum Immunol* 6(1): 45-52.

Breeden L, Nasmyth K. (1985) Regulation of the yeast HO gene. *Cold Spring Harbor Quant Biol* 50: 643-650.

Burbelo PD, Hall A. (1995) Hot numbers in signal transduction. *Current Biology* 5(2): 95-6.

Burgoyne RD, Cheek TR, Norman KM. (1986) Identification of a secretory granule-binding protein as caldesmon. *Nature* 319: 68-70.

Burgoyne RD, Morgan A, Robinson I, Pender N, Cheek T. (1993) Exocytosis in adrenal chromaffin cells. *J Anat* 183:309-14.

Burkhardt JK, Wiebel FA, Hester S, Argon Y. (1993) The giant organelles in beige and Chediak-Higashi fibroblasts are derived from late endosomes and mature lysosomes. *J Exp Med.* 178(6): 1845-56.

Cairo MS, Vandeven C, Toy C, Tischler D, Sender L. (1988) Fluorescent cytometric analysis of polymorphonuclear leukocytes in Chediak-Higashi syndrome: diminished C3bi receptor expression (OKM1) with normal granular cell density. *Pediatr Res* 24(6):673-6.

Chamberlain LH, Roth D, Morgan A, Burgoyne RD. (1995) Distinct effects of α-SNAP, 14-3-3 proteins and calmodulin on priming and triggering of regulated exocytosis. *J Cell Biol* 130: 1063-70.

Chediak M. (1952) Nouvelle anomalie leucocytaire de caractere constitutionnel et familial. *Rev Hematol* 7, 362.

Church GM, Gilbert W. (1984) Genomic sequencing. *Proc Natl Acad Sci USA*, 81, 1991-1995.

Clark EA, Roths JB, Murphy ED, Ledbetter JA, Clagett JA. (1982) The beige (bg) gene influences the development of autoimmune disease in SB/Le male mice. In: NK cells and other natural effector cells. Heberman-RB (ed). Academic Press, p. 301-306.

Clawson CC, White JG, Repine JE. (1978) The Chediak-Higashi syndrome. Evidence that defective leukotaxis is primarily due to an impediment by giant granules. *Am J Pathol* 92(3): 745-53.

Colgan SP, Blancquaert AM, Thrall MA, Bruyninckx WJ. (1992) Defective in vitro motility of polymorphonuclear leukocytes of homozygote and heterozygote Chediak-Higashi cats. *Vet Immunol Immunopoathol* 31(3-4): 205-27.

Colgan SP, Gasper PW, Thrall MA, Boone TC, Blancquaert AM, Bruyninckx WJ. (1992) Neutrophil function in normal and Chediak-Higashi syndrome cats following administration of recombinant canine granulocyte colony-stimulating factor. *Exp Hematol* 20(10):1229-34.

Colgan SP, Hull-Thrall MA, Gasper PW. (1989) Platelet aggregation and ATP secretion in whole blood of normal cats and cats homozygous and heterozygous for Chediak-Higashi syndrome. *Blood Cells* 15(3): 585-95.

Davletov B, Sontag JM, Hata Y, Petrenko AG, Fyks EM, Jahn R, Suedhof TC. (1993) Phosphorylation of synaptotagmin I by casein kinase II. *J Biol Chem* 268:6816-22.

DeLorenzo RJ. (1981) The calmodulin hypothesis of neurotransmission. *Cell Calcium* 2: 365-85.

Devereux J, Haeberli M, Smithies O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res*, 12, 387-395.

Fu H, Xia K, Pallas DC, Cui C, Conroy K, Narsimhan RP, Mamon H, Collier RJ, Roberts TM. (1994) Interaction of the protein kinase Raf-1 with 14-3-3 proteins. *Science* 266: 126-29.

Fujita Y, Fukui K, Kotani H, Kimura T, Hata Y, Sudhof TC, Scheller RH, Takai Y. (1996) *J Biol Chem* 271: 7265-68.

Fukai K, Oh J, Karim MA, Moore KJ, Kandil HH, Ito H, Buerger J, Spritz RA. (1996) Homozygosity mapping of the gene for Chediak-Higashi syndrome to chromosome 1q42-

q44 in a segment of conserved synteny that includes the mouse *beige* locus (*bg*). *Am J Hum Genet* 59:620-624.

Gallin JI, Bujak JS, Patten E, Wolff SM. (1974) Granulocyte function in the Chediak-Higashi syndrome of mice. *Blood* 43(2): 201-6.

Gallin JI, Klimerman JA, Padgett GA, Wolff SM. (1975) Defective mononuclear leukocyte chemotaxis in the Chediak-Higashi syndrome of humans, mink, and cattle. *Blood* 45(6): 863-70.

Gish W, States DJ. (1993) Identification of protein coding regions by database similarity search. *Nat Genet* 3(3):266-272.

Gong W, Emanuel BS, Collins J, Kim DH, Wang Z, Chen F, Zhang G, Roe B, Budarf M. (1996) A transcription map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. *Hum Mol Genet* 5: 789-800.

Goodrich KH, Holcombe RF. (1995) Genetic localization of the gene for Chediak-Higashi syndrome to human chromosome 1q and linkage to *nidogen*. *J Investig Med* 43(1), Suppl.1, 13A.

Green EL. (1981) Linkage, recombination and mapping. In "Genetics and Probabilities in Animal Breeding Experiments," pp. 77-113, Macmillan, New York.

Grossi CE, Crist WM, Abo T, Velardi A, Cooper MD. (1985) Expression of the Chediak-Higashi lysosomal abnormality in human peripheral blood lymphocyte subpopulations. *Blood* 65(4): 837-44.

Guo H, Sekiguchi M, Shimai K. (1992) Cytoarchitectonic abnormalities in cerebellum and hipocampal formation of beige mutant mouse. *Tokai J Exp Clin Med* 17(1), 53-61.

Hamanaka SC, Gilbert CS, White DA, Parmley RT. (1993) Ultrastructural morphology, cytochemistry, and morphometry of eosinophil granules in Chediak-Higashi syndrome. *Am J Pathol* 143(2): 618-27.

Hayakawa H, Kobayashi N, Yata J. (1986) Primary immunodeficiency diseases and malignancy in Japan. *Jpn J Cancer Res*. 77(1), 74-9.

Henikoff S, Henikoff JG. (1991) Automated assembly of protein blocks for database searching. *Nucleic Acids Research* 19 (23):6565-6572.

Higashi O. (1954) Congenital gigantism of peroxidase granules. The first case ever reported of qualitative abnormality of peroxidase. *The Tohoku J Exp Med* 59, 315.

Hillier L, Clark N, Dubuque T, Elliston K, Hawkins M, Holman M, Hultman M, Kucaba T, Le M, Lennon G, Marra M, Parsons J, Rifkin L, Rohlfing T, Soares M, Tan F, Trevaskis E, Waterson R, Williamson A, Wohldmann P, Wilson R. (1995, 1997) WashU-Merck EST Project.

Hirling H, Scheller RH. (1996) Phosphorylation of synaptic vesicle proteins: modulation of the αSNAP interaction with the core complex. *Proc Natl Acad Sci* 93: 11945-49.

Holcombe RF. (1992) Interleukin-2-induced cytotoxicity of Chediak-Higashi lymphocytes. *Acta Hematol* 87(1-2):45-8.

Holcombe RF, Jones KL, Stewart RM. (1994) Lysosomal enzyme activities in Chediak-Higashi syndrome: evaluation of lymphoblastoid cell lines and review of the literature. *Immunudeficiency* 5, 131-140.

Holcombe RF, Stephenson DA, Zweidler A, Stewart RM, Chapman VM, Seidman JG. (1991) Linkage of loci associated with two pigment mutations on mouse Chromosome 13. *Genet Res* 58, 41-50.

Holcombe RF, Strauss W, Owen FL, Boxer LA, Warren RW, Conley ME, Ferrara J, Leavitt RY, Fauci AS, Taylor BA, Seidman JG. (1987) Relationship of the genes for Chediak-Higashi syndrome (beige) and the T-cell receptor γ-chain in mouse and man. *Genomics* 1:287-91.

Holcombe RF, van de Griend R, Ang SL, Bolhuis RL, Seidman JG. (1990) Gamma-delta T cells in Chediak-Higashi syndrome. *Acta Hematol* 83(4):193-7.

Ishimura T, Isobe T, Okuyama T, Yamauchi T, Fujisawa H. (1987) Brain 14-3-3 protein is an activator protein that activates tryptophan 5-monooxygenase and tyrosine 3-monooxygenase in the presence of Ca^{2+}, calmodulin-dependent protein kinase II. *FEBS Lett* 219:79-82.

Ishino M, Ohba T, Inazawa J, Sasaki H, Ariyama Y, Sasaki T. (1997) Identification of an Efs isoform that lacks the SH3 domain and chromosomal mapping of human Efs. *Oncogene* 15: 1741-45.

Isobe T, Hiyane Y, Ishimura T, Okuyama T, Takahashi N, Nakajo S, Nakaya K. (1992) Activation of protein kinase C by the 14-3-3 proteins homologous with Exo1 protein that stimulates calcium-dependent exocytosis. *FEBS Lett* 308:121-24.

Ito H, Fukuda Y, Murata K, Kimura A. (1983) Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* 153(1):163-168.

Ito M, Sato A, Tanabe F, Ishida E, Takami Y, Shigeta S. (1989) The thiol proteinase inhibitors improve the abnormal rapid down-regulation of protein kinase C and the impaired natural killer cell activity in (Chediak-Higashi syndrome) beige mouse. *Biochem Biophys Res Commun* 160: 433-40.

Ito M, Tanabe F, Takami Y, Sato A, Shigeta S. (1988) Rapid down-regulation of protein kinase C in (Chediak-Higashi syndrome) beige mouse by phorbol ester. *Biochem Biophys Res Commun* 153: 648-56.

Jenkins NA, Justice MA, Gilbert DJ, Chu ML, Copeland NG. (1991) Nidogen/entactin (*Nid*) maps to the proximal end of mouse chromosome 13 linked to beige (*bg*) and identifies a new region of homology between mouse and human chromosomes. *Genomics* 9:401-3.

Justice MJ, Silan CM, Ceci JD, Buchberg AM, Copeland NG, Jenkins NA. (1990) A molecular genetic linkage map of mouse chromosome 13 anchored by the beige (*bg*) and satin (*sa*) loci. *Genomics* 6:341-351.

Kaerre K, Klein GO, Kiessling R, Klein G, Roder JC. (1980) Low natural in vivo resistance to syngeneic leukaemias in natural killer-deficient mice. *Nature* 284(5757): 624-6.

Kaiser C, Michaelis S, Mitchell A. (1994) Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, New York, p.209.

Karim MA, Nagle DL, Kandil HH, Buerger J, Moore KJ, Spritz RA. (1997) Mutations in the Chediak-Higashi syndrome gene (CHS1) indicate requirement for the complete 3801 amino acid CHS protein. *Hum Mol Genet* 6: 1087-89.

Kingsmore SF, Barbosa MDFS, Nguyen QA, Ashley JA, Blaydes SM, Tchernev VT, Detter CD, Lovett M. (1996b) Physical mapping of the beige critical region on mouse chromosome 13. *Mamm Genome* 7:773-5.

Kingsmore SF, Barbosa MDFS, Tchernev VT, Detter CD, Lossie AC, Seldin MF, Holcmbe RF. (1996a) Positional cloning of the Chediak-Higashi syndrome gene: genetic mapping of the beige locus on mouse chromosome 13. *J Invest Med* 44(8):454-61.

Kinugawa N, Ohtani T. (1985) Beneficial effects of high-dose intravenous gammaglobulin on the accelerated phase of Chediak-Higashi syndrome. *Helv Paediatr Acta* 40(2-3):169-172.

Knight SP, Richardson MM, Osmand AP, Stakkestad A, Potter NT. (1997) Expression and distribution of the dentatorubral-pallidoluysian atrophy gene product (atrophin 1/drplap) in neuronal and non-neuronal tissues. *J Neurol Sci* 146: 19-26.

Koffer A, Tatham PER, Gomperts BD. (1990) Changes in the state of actin during the exocytotic reaction in mast cells. *J Cell Biol* 111: 919-27.

Komada M, Kitamura N. (1994) Regulatory role of major tyrosine autophosphorylation site of kinase domain of c-Met receptor (scatter factor/hepatocyte growth factor receptor. *J Biol Chem* 269: 16131-36.

Komada M, Kitamura N. (1995) Growth factor-induced tyrosine phosphorylation of Hrs, a novel 115-kilodalton protein with a structurally conserved putative zinc finger domain. *Mol Cell Biol* 15: 6213-21.

Komada M, Masaki R, Yamamoto A, Kitamura N. (1997) Hrs, a tyrosine kinase substrate with conserved double zinc finger domain, is localized to the cytoplasmic surface of early endosomes. *J Biol Chem* 272: 20538-44.

Kondo N, Shimozawa N, Asano J, Imamura A, Orii T. (1994) Chediak-Higashi syndrome with cerebellar cortical atrophy detected by MRI. *Clin Genet* 46(6), 439-40.

Kreimer DJ, Khotimchenko YS. (1995) Cytoplasm calcium-binding proteins of germ cellsand embryos of the sea urchin. *Comp Biochem Physiol* 110A: 95-105.

Kubo A, Sasada M, Nishimura T, Moriguchi T, Kakita T, Yamamoto K, Uchino H. (1987) Oxygen radical generation by polymorphonuclear leukocytes of beige mice. *Clin Exp Immunol* 70(3):658-63.

Lambright DG, Sondek J, Bohm A, Skiba NP, Hamm HE, Sigler PB. (1996) The 2 Å crystal structure of a heterotrimeric G protein. *Nature* 379:311-19.

Lane PW. (1971) *Xt-bg-cr* linkage. *Mouse News Lett* 45:29.

Laskey RA, Goerlich D, Madine MA, Makkerh JPS, Romanowski P. (1996) Regulatory roles of the nuclear envelope. *Exp Cell Res* 229: 204-11.

Lockman LA, Kennedy WR, White JG. (1967) The Chediak-Higashi syndrome:electrophysiological and electron microscopic observations on the peripheral neuropathy. *J Pediatr* 70, 942-51.

Lyon MF, Meredith R. (1969) Muted, a new mutant affecting coat colour and otoliths of the mouse, and its position in linkage group XIV. *Genet Res* 14:163-66.

Mahoney KH, Morse SS, Morahan PS. (1980) Macrophage functions in beige (Chediak-Higashi syndrome) mice. *Cancer Res* 40(11): 3934-9.

Margolis RL, Li SH, Young WS, Wagster MV, Stine OC, Kidwai AS, Ashworth RG, Ross CA. (1996) DRPLA gene (atrophin 1) sequence and mRNA expression in human brain. *Mol Brain Res* 36: 219-26.

Marklund U, Larsson N, Gradin HM, Brattsand G, Gullberg M. (1996) Oncoprotein 18 is a phosphorylation-responsive regulator of microtubule dynamics. *EMBO J* 15:5290-98.

Martin H, Rostas J, Patel Y, Aitken A. (1994) Analysis of subcellular distribution of 14-3-3 isoforms in rat brain using specific antibodies. *J Neurochem* 63:2259-65.

Matsuda M, Okabe T, Sugimoto N, Senda T, Fujita H. (1994) Tetanus toxin and Clostridium perfringens enterotoxin as tools for the study of exocytosis. *Ann N Y Acad Sci* 710: 94-106.

Matteoni R, Kreis TE. (1987) Translocation and clustering of endosomes and lysosomes depends on microtubules. *J Cell Biol* 105:1253-65.

Menard M, Meyers KM. (1988) Storage pool deficiency in cattle with Chediak-Higashi syndrome results from an absence of dense granule precursors in their megakaryocytes. *Blood* 72(5): 1726-34.

Merino F, Klein GO, Henle W, Ramirez-Duque P, Forsgren M, Amesty C. (1983) Elevated antibody titers to Epstein-Barr virus and low natural killer cell activity in patients with Chediak-Higashi syndrome. *Clin Immunol Immunopathol* 27(3):326-39.

Meyers KM, Seachord CL, Benson K, Fukami M, Holmsen H. (1983) Serotonin accumulation in granules of storage pool-deficient platelets of Chediak-Higashi cattle. *Am J Physiol* 245(1): H150-8.

Meyers KM, Stevens DR, Padgett GA. (1974) A platelet serotonin anomaly in the Chediak-Higashi syndrome. *Res Commun Chem Pathol Pharmacol.* 7(2), 375-80.

Misra VP, King RH, Harding AE, Muddle JR, Thomas PK. (1991) Peripheral neuropathy in the Chediak-Higashi syndrome. *Acta Neuropathol Berl.* 81(3), 354-8.

Mitchell PJ, Tjian R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245 (4916): 371-8.

Morrison D. (1994) 14-3-3: modulators of signaling proteins? *Science* 266: 56-7.

Morrison D. (1995) Mechanisms regulating Raf1 activity in signal transduction pathways. *Mol Reprod Dev* 42: 507-14.

Moskowitz N, Schook W, Lisanto M, Hua E, Puszkin S. (1982) Calmodulin affinity for brain coated vesicle protein. *J Neurochem* 38: 1742.

Murphy ED, Roths JB. (1978) Purkinje cell degeneration, a late effect of beige mutations in mice. *Jackson Lab Ann Rep* 49, 108-9.

Myers SM, Eng C, Ponder BAJ, Mulligan LM. (1995) Characterization of *RET* proto-oncogene 3' splicing variants and polyadenylation sites: a novel C-terminus for RET. *Oncogene* 11:2039-45.

Nagase T, Seki N, Ishikawa K, Tanaka A, Nomura N. (1996) Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (KIAA0161 – KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* 3 (1) :17-24.

Nagle DL, Karim MA, Woolf EA, Holmgren L, Bork P, Misumi DJ, McGrail SH, Dussault BJ, Perou CM, Boissy RE, Duyk GM, Spritz RA, Moore KJ. (1996) Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome. *Nat Genet* 14:307-11.

Nakai K, Kanehisa M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14:897-911.

Novak EK, McGarry MP, Swank RT. (1985) Correction of symptoms of platelet storage pool deficiency in animal models for Chediak-Higashi syndrome and Hermansky-Pudlak syndrome. *Blood* 66(5): 1196-201.

Oka JA, Weigel PH. (1983) Microtubule-depolymerizing agents inhibit asialo-orosomucoid delivery to lysosomes but not its endocytosis or degradation in isolated rat hepatocytes. *Biochim Biophys Acta* 763:368-76.

Oliver JM, Zurier RB, Berlin RD. (1975) Concavalin A cap formation on polymorphonuclear leukocytes of normal and beige (Chediak-Higashi) mice. *Nature* 253:471-73.

Olsen DR, Nagayoshi T, Fazio M, Mattei M-G, Passage E, Weil D, Timpl R, Chu M-L, Uitto J. (1989) Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to chromosome 1q43. *Amer J Hum Genet* 44:876-885.

Owen FL, Taylor BA, Zweidler A, Seidman JG. (1986) The murine γ-chain of the T cell receptor is closely linked to a spermatocyte specific histone gene and the beige coat color locus on chromosome 13. *J Immunology* 137:1044-46.

Peifer M, Berg S, Reynolds A. (1994) A repeating amino-acid motif shared by proteins with diverse cellular functions. *Cell* 76: 789-91.

Penner JD, Prieur DJ. (1987) Interspecific genetic complementation analysis with fibroblasts from humans and four species of animals with Chediak-Higashi syndrome. *Am J Med Genet*. 28(2), 455-70.

Perin MS, Brose N, Jahn R, Suedhof TC. (1991) *J Biol Chem* 266:623-29.

Perou C, Justice MJ, Pryor RJ, Kaplan J. (1996a) Complementation of the beige mutation in cultured cells by episomally replicating murine yeast artificial chromosomes. *Proc Natl Acad Sci USA* 93: 5905-09.

Perou CM, Kaplan J. (1993) Chediak-Higashi syndrome is not due to a defect in microtubule-based lysosomal mobility. *J Cell Sci*. 106 (Pt 1), 99-107.

Perou CM, Moore KJ, Nagle DL, Misumi DJ, Woolf EA, McGrail SH, Holmgren L, Brody BJD, Monroe CA, Duyk GM, Pryor RJ, Li L, Justice M, Kaplan J. (1996b) Identification of the murine beige gene by YAC complementation and positional cloning. *Nature Genetics* 13, 303-308.

Perrin D, Moeller K, Hanke K, Soeling HD. (1992) cAMP and Ca^{2+} mediated secretion in parotid acinar cells is associated with reversible changes in the organization of the cytoskeleton. *J Cell Biol* 116: 127-34.

Pettit RE, Berdal KG. (1984) Chediak-Higashi syndrome. Neurologic appearance. *Arch Neurol* 41(9), 1001-2.

Pezeshkpour G, Kurent JS, Krarup C, Buchthal F, Fauci AS. (1986) Peripheral neuropathy in Chediak-Higashi syndrome. *J Neuropathol Exp Neurol* 45: 353.

Pflumio F, Fonteneau P, Loor F. (1990) Impaired antibody response of C57BL/6 beige mutant mice to a thymus-independent type 2 antigen. *Immunol Lett* 23(4): 269-74.

Phillips LL, Kaplan HA, Padgett GA, Gorham JR. (1967) Comparative studies on the Chediak-Higashi syndrome: coagulation and fibrinolytic mechanisms of mink and cattle. *Am J Vet Clin Path* 1: 1-6.

Pratt HL, Carroll RC, Jones JB, Lothrop CD Jr. (1991) Platelet aggregation, storage pool deficiency, and protein phosphorylation in mice with Chediak-Higashi syndrome. *Am J Vet Res* 52(6): 945-50.

Ramirez-Duque P, Arends T, Merino F. (1982) Chediak-Higashi syndrome: description of a cluster in a Venezuelan-Andean isolated region. *J Medicine*, 13 (5-6), 431-51.

Rendu F, Breton-Gorius J, Lebret M, Klebanoff C, Buriot D, Griscelli C, Levy-Toledano S, Caen JP. (1983) Evidence that abnormal platelet functions in human Chediak-Higashi syndrome are the result of a lack of dense bodies. *Am J Pathol* 111(3): 307-14.

Reuther GW, Fu H, Cripe LD, Collier RJ, Pendergast AM. (1994) Association of the protein kinases c-Bcr and Bcr-Abl with proteins of the 14-3-3 family. *Science* 266: 129-133.

Roder J, Duwe A. (1979) The beige mutation in the mouse selectively impairs natural killer cell function. *Nature* 278(5703): 451-3.

Roder JC, Haliotis T, Laing L, Kozbor D, Rubin P, Pross H, Boxer LA, White JG, Fauci AS, Mostowski H, Matheson DS. (1982) Further studies of natural killer cell function in Chediak-Higashi patients. *Immunology* 46(3): 555-60.

Roder JC, Lohmann-Matthes ML, Domzig W, Wigzell H. (1979) The beige mutation in the mouse. II. Selectivity of the natural killer (NK) cell defect. *J Immunol* 123(5): 2174-81.

Roder JC, Todd RF, Rubin P, Haliotis T, Helfand SL, Werkmeister J, Pross HF, Boxer LA, Schlossman SF, Fauci AS. (1983) The Chediak-Higashi gene in humans. III. Studies on the mechanisms of NK impairment. *Clin Exp Immunol* 51(2): 359-68.

Roep BO, Arden SD, de Vries RRP, Hutton JC. (1990) T-cell clones from a type 1 diabetes patient respond to insulin secretory granule proteins. *Nature* 345: 632-34.

Ross CA, Becher MW, Colomer V, Engelender S, Wood JD, Sharp AH. (1997) Huntington's disease and dentatorubral-pallidoluysian atrophy: proteins, pathogenesis and pathology. *Brain Pathol* 7: 1003-16.

Roth D, Burgoyne RD. (1995) Stimulation of catecholamine secretion from adrenal chromaffin cells by 14-3-3 proteins is due to reorganisation of the cortical actin network. *FEBS Lett* 374: 77-81.

Roth D, Morgan A, Martin H, Jones D, Martens GJM, Aitken A, Burgoyne RD. (1994) Characterization of 14-3-3 proteins in adrenal chromaffin cells and demonstration of isoform-specific phospholipid binding. *Biochem J* 301: 305-10.

Sambrook J, Fritsch EF, Maniatis T. (1989) In "Molecular cloning: a laboratory manual", Vol. 2, p. 9.16-9.23, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sato A. (1955) Chediak and Higashi's disease: probable identity of "a new leucocytal anomaly (Chediak)" and "congenital gigantism of peroxidase granules (Higashi)". *The Tohoku J Exp Med*, 61, 201.

Sato A, Tanabe F, Ito M, Ishida E, Shigeta S. (1990) Thiol proteinase inhibitors reverse the increased protein kinase C down-regulation and concavalin A cap formation in

polymorphonuclear leukocytes from Chediak-Higashi syndrome (beige) mouse. *J Leukoc Biol* 48: 377-81.

Schiebler W, Jahn R, Doucet JP, Rothlein J, Greengard P. (1986) Characterization of synapsin I binding to small synaptic vesicles. *J Biol Chem* 261: 8383-90.

Sherman F. (1991) Getting started with yeast. *Methods Enzymol* 194: 3-21.

Shinozaki K, Maruyama K, Kume H, Kuzume H, Obata K. (1997) A novel brain gene, norbin, induced by treatment of tetraethylammonium in rat hippocampal slice and accompanied with neurite-outgrowth in neuro 2a cells. *Biochem Biophys Res Commun*, 240 (3): 766-71.

Sjaastad OV, Blom AK, Stormorken H, Nes N. (1990) Adenine nucleotides, serotonin, and aggregation properties of blue foxes (Alopex lagopus) with the Chediak-Higashi syndrome. *Am J Med Genet* 35(3): 373-8.

Sobel A. (1991) Stathmin: a relay phosphoprotein for multiple signal transduction? *Trends Biochem Sci* 16:301-05.

St. Amand W, Cupp MB. (1958) New linkage. *Mouse News Lett* 19:38.

Steinbrinck W. (1948) Ueber eine neue Granulationsanomalie der Leukocyten. *Deutsch Archiv Klin Med*, 193, 577.

Sugimoto Y, Kusakabe T, Kai T. (1995) Analysis of the *in vitro* translation product of a novel-type *Drosophila melanogaster* aldolase mRNA in which two carboxy-terminal exons remain unspliced. *Arch Biochem Biophys* 323: 361-66.

Sung JH, Meyers JP, Stadlan EM, Cowen D, Wolf A. (1969) Neuropathological changes in Chediak-Higashi disease. *J Neuropathol Exp Neurol* 28, 86-118.

Swanson JA, Bushnell A, Silverstein SC. (1987) Tubular lysosome morphology and distribution within macrophages depend upon the integrity of cytoplasmic microtubules. *Proc Natl Acad Sci USA* 84:1921-25.

Swanson JA, Locke A, Ansel P, Hollenbeck PJ. (1992) Radial movement of lysosomes along microtubules in permealized macrophages. *J Cell Sci* 103:210-9.

Sygiyama T, Nishio Y, Kishimoto T, Akira S. (1996) Identification of alternative splicing form of Stat2. *FEBS Lett* 381:191-94.

Talmage JE, Meyers KM, Prieur DJ, Starkey JR. (1980) Role of NK cells in tumour growth and metastasis in beige mice. *Nature* 284:622-4.

Targan SR, Oseas R. (1983) The "lazy" NK cells of Chediak-Higashi syndrome. *J Immunol* 130(6): 2671-4.

Tchernev VT, Barbosa MDFS, Detter JC, Patel TD, Achey K, Wakeland EK, Gueorguieva RV, Yang MCK, Gossler A, Kingsmore SF (1997). Genetic mapping of 20 novel expressed sequence tags from midgestation mouse embryos suggests chromosomal clustering. *Genomics* 40, 170-174.

Toker A, Ellis CA, Sellers LA, Aitken A. (1990) Protein kinase C inhibitor proteins. Purification from sheep brain and sequence similarity to lipocortin and 14-3-3 protein. *Eur J Biochem* 191: 421-29.

Toker A, Sellers LA, Amess B, Patel Y, Harris A, Aitken A. Multiple isoforms of a protein kinase C inhibitor (KCIP-1/14-3-3) from sheep brain. Amino acid sequence of phosphorylated forms. (1992) *Eur J Biochem* 206: 453-461.

Trapp T, Holsboer F. (1996) Nuclear orphan receptor as a repressor of glucocorticoid receptor transcriptional activity. *J Biol Chem* 17: 9879-82.

Trifaro JM, Vitale ML, Del Castillo AR. (1992) Cytoskeleton and molecular mechanisms in neurotransmitter release by neurosecretory cells. *Europ J Pharmacol* 225: 83-104.

Uyama E, Hirano T, Ito K, Nakashima H, Sugimoto M, Naito M, Uchino M, Ando M. (1994) Adult Chediak-Higashi syndrome presenting as parkinsonism and dementia. *Acta Neurol Scand* 89(3), 175-83.

Van De Wetering M, Castrop J, Korlinev V, Clevers H. (1996) Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. *Mol Cell Biol* 16:745-752.

Virelizier JL, Griscelli C. (1980) Interferon administration as an immunoregulatory and antimicrobial treatment in children with defective interferon secretion. *Primary immunodeficiencies*. Elsevier/North Holland Biomedical Press, 1980. Pp. 473-84.

Wallace JC, Henikoff S. (1992) PATMAT: a searching and extraction program for sequence, pattern and block queries and databases. *Comput Appl Biosci* 8(3):249-254.

Watkins DT, Cooperstein SJ. (1983) Role of calcium and calmodulin in the interaction between islet cell secretion granules and plasma membranes. *Endocrinology* 112: 766.

Wiese S, Murphy DB, Schlung A, Burfeind P, Schmundt D, Schnuelle V, Mattei MG, Thies U. (1995) The genes for human brain factor 1 and 2, members of the fork head gene family, are clustered on chromosome 14q. *Biochim Biophys Acta* 1262: 105-112.

Willingham MC, Spicer SS, Vincent RA Jr. (1981) The origin and fate of large dense bodies in beige mouse fibroblasts. Lysosomal fusion and exocytosis. *Exp Cell Res* 136(1), 157-68.

Windhorst DB, Zelickson AS, Good RA. (1966) Chediak-Higashi syndrome: hereditary gigantism of cytoplasmic organelles. *Science*, 151, 81-83.

Windhorst DB, Zelickson AS, Good RA. (1968) A human pigmentary dilution based on a heritable subcellular structural defect--the Chediak-Higashi syndrome. *J Invest Dermatol.* 50(1), 9-18.

Yakel JL. (1997) Calcineurin regulation of synaptic function: from ion channels to transmitter release and gene transcription. *Trends Pharmacol Sci* 18: 124-34.

Yamashita T, Wu N, Kupfer G, Corless C, Joenje H, Grompe M, D'Andrea AD. (1996) Clinical variability of Fanconi anemia (Type C) results from expression of an amino terminal truncated Fanconi anemia complementation group C polypeptide with partial activity. *Blood* 87:4424-32.

Yang N, Shigeta H, Shi H, Teng CT. (1996) Estrogen-related receptor, hERRl, modulates estrogen receptor-mediated response of human lactoferrin gene promoter. *J Biol Chem* 271: 5795-804.

Yazawa I, Nukina N, Hashida H, Goto J, Yamada M, Kanazawa I. (1995) Abnormal gene product identified in hereditary dentatorubral-pallidoluysian atrophy (DRPLA) brain. *Nat Genet* 10: 99-103.

Yegin O, Sanal O, Yeralan O, Gurgey A, Berkel AI. (1983) Defective lymphocyte locomotion in Chediak-Higashi syndrome. *Am J Dis Child* 137(8):771-3.

Zhao H, Boissy YL, Zalfa AM, King RA, Nordlund JJ, Boissy RE. (1994) On the analysis of the pathophysiology of Chediak-Higashi syndrome. Defects expressed by cultured melanocytes. *Lab Investigation*, 71, 25-34.

Zhao W, Manlley JL. (1996) Complex alternative mRNA processing generates an unexpected diversity of poly(A) polymerase forms. *Mol Cell Biol* 16:2378-2386.

Zupan LA, Steffens DL, Berry ML, Gross RW. (1992) *J Biol Chem* 267:8707-10.

BIOGRAPHICAL SKETCH

Velizar T. Tchernev was born in Sofia, Bulgaria. In 1987, he graduated from high school with extensive study of the German language in Sofia. Velizar Tchernev received his Doctor of Medicine degree from the Higher Medical Institute-Sofia, Bulgaria, in 1993. He also obtained diplomas for Translator of Medical Literature from German (1990) and from English (1991) languages from the Higher Medical Institute-Sofia. In 1994, Velizar Tchernev entered the Ph.D. program in Immunology and Molecular Pathology at the Department of Pathology and Laboratory Medicine, College of Medicine, University of Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Stephen F. Kingsmore, Chair
Assistant Professor of Pathology,
Immunology and Laboratory Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Margaret R. Wallace
Associate Professor of Pediatrics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Eric S. Sobel
Associate Professor of Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Michael R. Bubb
Assistant Professor of Medicine

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Edward K. Wakeland
Professor of Pathology, Immunology and Laboratory Medicine

This dissertation was submitted to the Graduate Faculty of the College of Medicine and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1998

Dean, College of Medicine

Dean, Graduate School

UNIVERSITY OF FLORIDA
3 1262 08555 3021